Madre de Dios (Peru): Hydrology, Sediment Transport, and Gold Mining

by

Wenjie Wang

Department of Civil and Environmental Engineering
Duke University

Date:_____

Approved:

_____
Amilcare Porporato, Supervisor

_____
Marco Marani

_____
Mukesh Kumar

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School of
Duke University

2014

ABSTRACT

Madre de Dios (Peru): Hydrology, Sediment Transport, and Gold Mining

by

Wenjie Wang

Department of Civil and Environmental Engineering
Duke University

Date:_____

Approved:

_____
Amilcare Porporato, Supervisor

_____
Marco Marani

_____
Mukesh Kumar

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Department of
Civil and Environmental Engineering in the Graduate School of
Duke University

2014

Abstract

Legal and illegal gold mining in the Madre de Dios area cause several environmental concerns linked with mercury pollution and deforestation, which make hydrological and sediment transport processes central for any assessment of ongoing and future environmental impacts. This thesis addresses the problem of estimating flow and transport processes in the virtually ungauged Madre de Dios basin (Peru) to provide a first basis on which transport of mercury adsorbed on the sediment may be evaluated. The thesis first applies a recent method to derive the probability distribution of liquid and solid discharges to gauged basins in the Tropical Zone (Puerto Rico). This application allows the evaluation of the uncertainty associated with estimates of flow and transport pdf's. Once the uncertainty is quantified, the same tool is applied to provide the first evaluation of the frequency distributions of discharge and sediment fluxes in the Madre de Dios area.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my deep gratitude to Professor Marani, my research advisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. I would also like to thank Gianluca Botter and Stefano Basso, for their advice and assistance in keeping my progress on schedule.

1. Introduction

1.1 Background

Gold is an important resource, which plays an irreplaceable role in many areas. In times of economic crisis global gold prices often rise, as they have over the past several years. As a response, gold mining has intensified in many developing countries [1, 2].

However, gold mining is associated with numerous environmental impacts, especially in the remote regions of the Peruvian Amazon, where it is associated with 1) deforestation, and 2) mercury pollution. In fact, the Food and Agriculture Organization (FAO) reported that "the annual deforestation rate has increased considerably – from 0.14% (of the total national forest cover) between 2000 and 2005 to 0.22% between 2005 and 2010" in Peru [3].

Mercury has been used in gold mining for centuries to separate and collect gold from the soil substrate. The modern gold-mining industry has now replaced mercury with cyanide, which is less toxic, but mercury is still used to produce more than 70% of the annual gold production in Peru [4]. In fact, small-scale miners in Madre de Dios still tend to use mercury, because they do not have the necessary know how to replace it, and because mercury is much cheaper [5, 6]. According to the United Nations Environment Program, small-scale miners globally use up to 1,350 tons of mercury each year. About 45–50 tonnes of mercury are used each year in Madre de Dios alone [6].

Gold miners pour mercury mixed with sediments into rivers, where it can potentially be transported for long distances downstream. Mercury also enters the food web through benthic fish and can produce significant accumulations in the human body. “In 2009, Fernandez discovered that the most-consumed fish species in Madre de Dios, such as the mota (Calophysus macropterus) and doncella (Pseudoplatystoma fasciatum), had the highest levels of mercury.” [6] According to the World Health Organization, people consume 7.5 times the safe limit of mercury if they eat this fish two times per week. Finally, during the gold extraction process, part of the mercury evaporates into the atmosphere, where it is transported and eventually returned to the ground over large areas [7]. The impacts of this atmospheric contamination pathway remain largely unexplored in the Madre de Dios area.

Because mercury is easily adsorbed onto sediment particles (particularly to the finer fraction), the evaluation of the impacts of mercury release into the environment must start from the development and application of models of water and sediment transport. The development of such models is, however, greatly hampered by the lack of hydrological observations. Streamflow and suspended and bedload transport observations are almost totally lacking in the Madre de Dios catchment. I address this problem here by using a recent approach to the estimation of residence times probability distributions (pdfs) developed by Botter et al, which requires a limit amount of information [8, 9]. Land use information is also scarce and rapidly evolving. Hence I

estimate land use distribution by use of remote sensing data, to substitute for the lacking field information.

In this framework, the objectives of the present thesis are twofold: 1) establish a methodology for the characterization of flow and transport processes in the MDD region, and 2) provide a first estimate of the amount of sediment transported across the MDD at the outlet of Puerto Maldonado, downstream of the main mining areas. This thesis starts by developing and applying a stochastic methodology for the evaluation of water and sediment fluxes to well gauged experimental catchments in Puerto Rico. This is intended to show the ability of the approach to capture the pdf's of liquid and solid discharge in a climate similar to that of the MDD. Subsequently, the methodology is applied to the MDD catchment to derive reference estimate of suspended sediment concentrations and total transport. The MDD does not have information to apply model as previously done. Therefore, the empirical parameters are used to estimate streamflow and sediment transports. The range of parameters is used to decide the uncertainty of results.

1.2 Material and methods

1.2.1 The Study area

1.2.1.1 Madre de Dios

The Madre de Dios region is located in southeastern Peru, bordering Brazil, Bolivia and the Peruvian regions of Puno, Cusco and Ucayali, in the Amazon Basin. Its capital is the city of Puerto Maldonado.

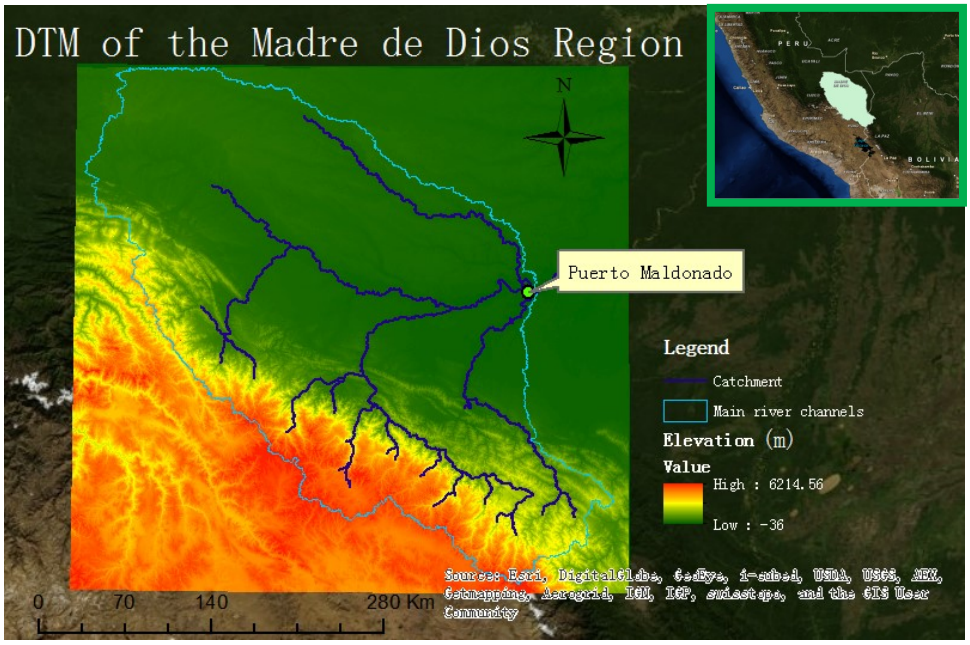
The study area is the Madre de Dios catchment, a part of the Amazon River watershed, with outlet at Puerto Maldonado (12°36'0"S 69°11'0"W). The total area of the catchment is about 85,000 km^2. The Highest elevation is 3,932 m and the lowest elevation is 183 m. The river's main tributaries include the Chivile, the Blue river, the White river, the Inambari, the Tambopata and the Colorado river [10]. According to the Harmonized World Soil Database (HWSD), the dominant soil types are silt loam and sandy loam [11].

Figure 1a shows the Madre de Dios catchment and the highest order channel in this area.

Figure 1b shows the DTM of the Madre de Dios catchment. The elevation data is provided by USGS, and I used it to extract the main river channels in the catchment.


UCAYALI
Madre de Dios Catchment
Cobija
PANDO
MADRE DE DIOS
Puerto Maldonado
CUZCO
Cusco
Abancay
APURÍMAC
Legend
Puerto Maldonado and rainfall station
Madre de Dios Region
Main river channels
N
0
87.5
175
350 Km
Juliaca
Lago Titicaca
PERU
BOLIVIA

(a)


DTM of the Madre de Dios Region
N
Puerto Maldonado
Legend
Catchment
Main river channels
Elevation (m)
Value
High : 6214.56
Low : -36
0
70
140
280 Km
Source: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community
PERU
BOLIVIA

(b)

Figure 1: a) Madre de Dios Catchment b) DTM of the Madre de Dios

According to data from the Gobierno Regional Madre de Dios, the average annual temperature is around 26 °C (79 °F) [max.: 34 °C (93 °F), min.: 21 °C (70 °F)]. The rainy season is from December to March, when torrential rainfall causes rivers to swell and often overflow their banks. Annual precipitation can be as much as 3 metres (9.8 ft) [12].

The rainfall data available for the study is from January 1998 to December 2008 (source: Peruvian National Water Authority - Autoridad Nacional del Agua). Table 7 and Figure 2 provide monthly rainfall in Puerto Maldonado [13].

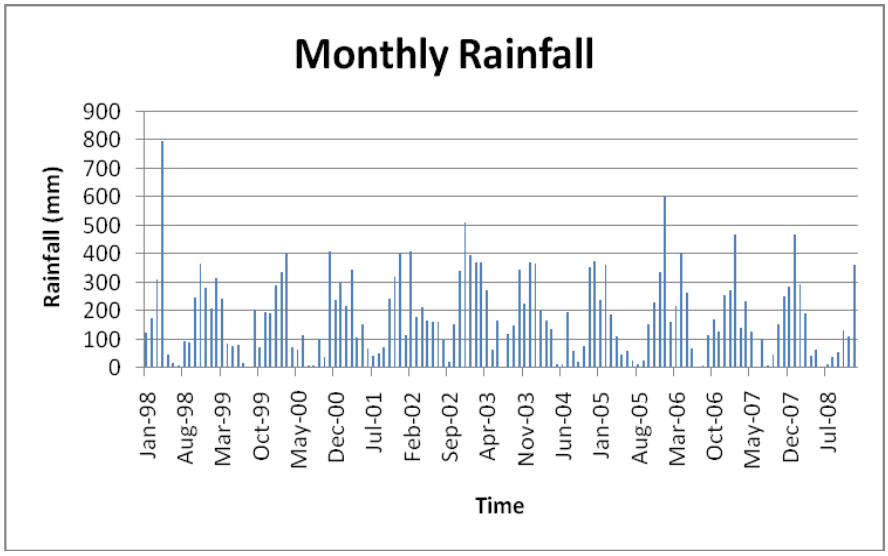


Figure 2 Monthly rainfall in Puerto Maldonado

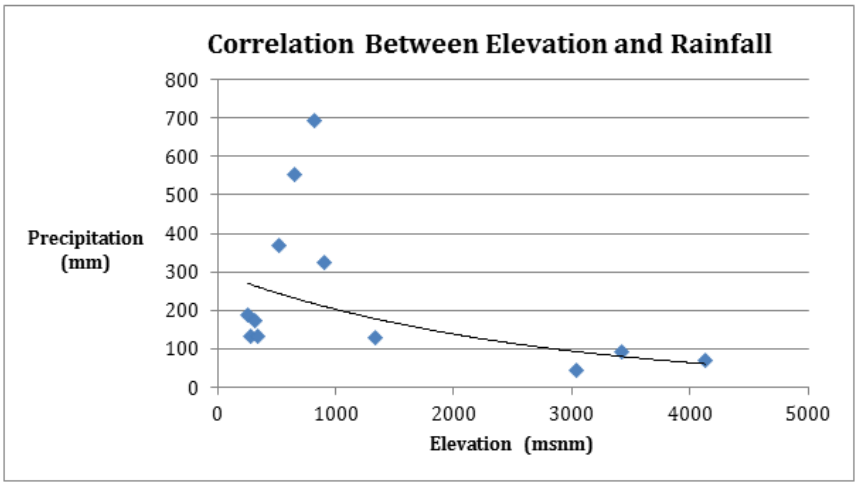


Figure 3: Scatterplot of rainfall and elevation, demonstrates a spike in rainfall at 820 meters above sea level. Provided by Axel Berky.

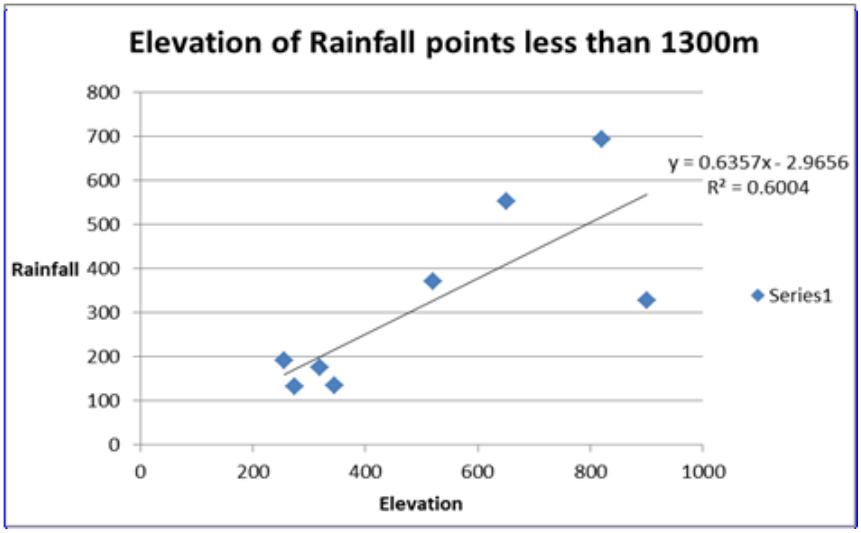


Figure 4: Scatterplot of rainfall and elevation, showing a high correlation between rainfall and elevation with an R^2 value of 0.6. Provided by Axel Berky.

Berky claims that "while orographic forcing certainly depends on multiple variables, rather than just on elevation, the scatterplot suggests that the more complex orographic effects in rainfall on the local regime can be surrogated by elevation alone, by assuming an increase in precipitation up to about 1,300 m and an approximately

constant mean at high elevations, above 1,300 m." [14] He separated the rain gauges into two groups by elevation, with those at elevations higher than 1,300 m in one group and rain gauges below 1,300m in another. "A linear regression was used for gauges below 1,300 m, while a constant mean was assumed for gauges above 1,300m. The residuals between rainfall observations with respect to the mean trends assumed above were then determined and interpolated using ordinary Kriging. The mean value for each location was then added to yield the final interpolated value. The two interpolated regions were then combined together to produce the final interpolated surface." [14] Due to the high spatial heterogeneity of rainfall distribution, associated with the presence of very high relief, rainfall is interpolated on the basis of a network of stations relatively well distributed throughout the catchment (Figure 5).

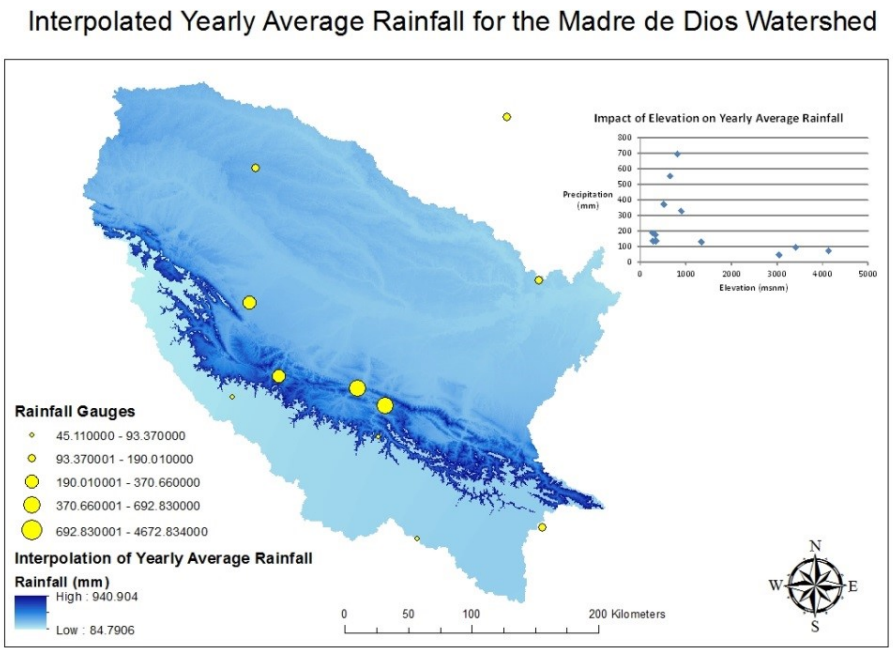


Figure 5: Rain gauges in Southeastern Peru and the resulting interpolated yearly rainfall for the Madre de Dios watershed. Provided by Axel Berky.

1.2.1.2 Puerto Rico experimental basins

USGS monitors several streams in Puerto Rico, acquiring rainfall, streamflow, and sediment transport observations. The Commonwealth of Puerto Rico has an area of 13,790 km^2, of which 8,870 km^2 is land and 4,921 km^2 is water [15]. The maximum length of the main island from east to west is 180 km, and the maximum width from north to south is 65 km [16]. Puerto Rico has 17 lakes, all artificial, and more than 50 rivers, most originating in the Cordillera Central [17]. Rivers in the northern region of the island are typically longer and with higher water flow rates than those of the south, since the south receives less rain than the central and northern regions [18].

Puerto Rico has an average temperature of 82.4 °F (28 °C) throughout the year, with an average minimum temperature of 66.9 °F (19 °C) and maximum of 85.4 °F (30 °C) [19]. One reason for choosing Puerto Rico as experimental area is the rainy season. Puerto Rico experiences the Atlantic hurricane season, similar to the remainder of the Caribbean Sea and North Atlantic oceans. On average, a quarter of its annual rainfall is contributed from tropical cyclones, which is more prevalent during periods of La Niña than El Niño. Autumn is the peak of the hurricane season, especially in September. Autumn months are very wet, with frequent and intense thunderstorms on the west coast [20].

I identified two catchments in Puerto Rico which are carefully monitored by the USGS. Stream station 1 (18°19'46"N, 65°45'04"W) is in the Rio Mameyes catchment, with a watershed area of about 18 km^2. Stream station 2 (18°13'53"N, 66°38'14"W) is in Rio Caonillas, with a watershed area of about 98 km^2.

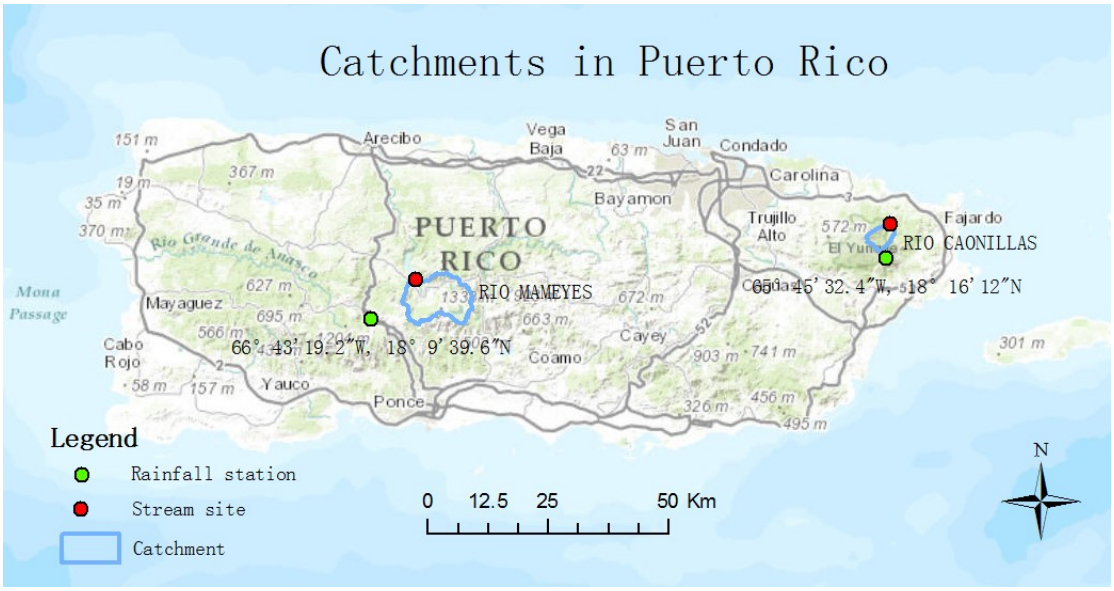


Figure 6: Two catchments in Puerto Rico. Stream station 1 is in Rio Mameyes and stream station 2 is in Rio Caonillas

Table 8, Figures 7, 8, 9 and 10 show monthly rainfall, discharge, daily sediment concentration and discharge data from December 1992 to August 2003 in Rio Mameyes, provided by the USGS Caribbean Water Science Center [21].

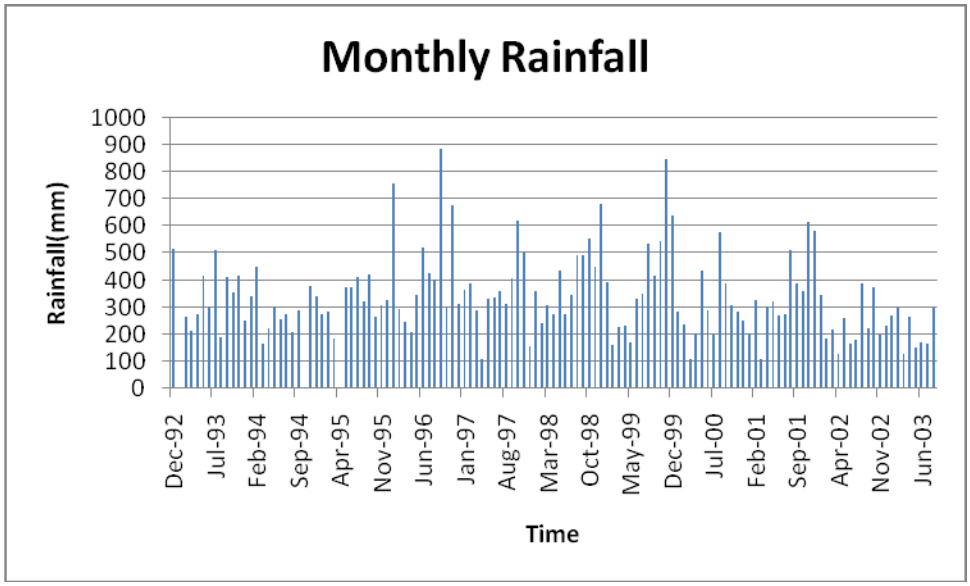


Figure 7: Monthly rainfall in Rio Mameyes

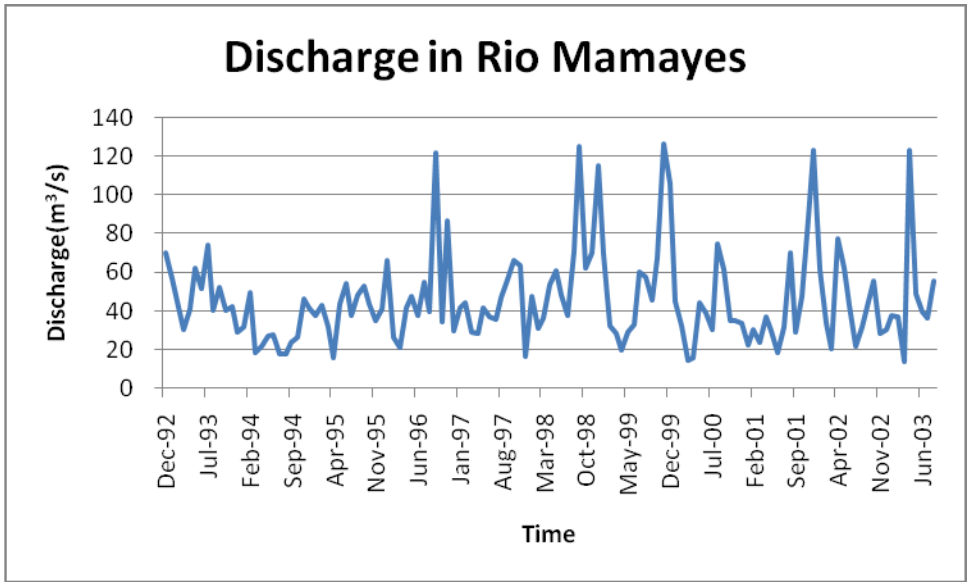


Figure 8: Monthly discharge in Rio Mameyes

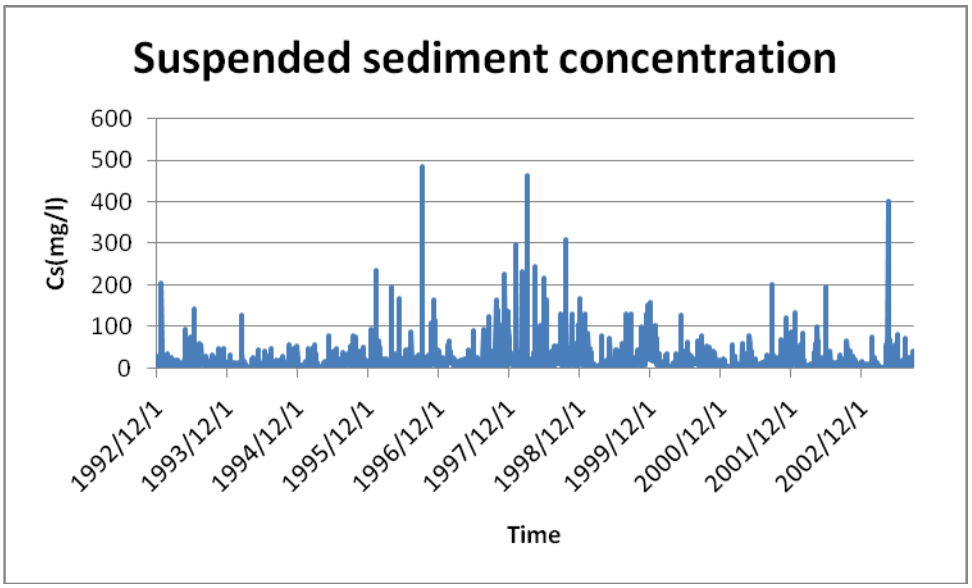


Figure 9: Daily suspended sediment concentration in Rio Mameyes

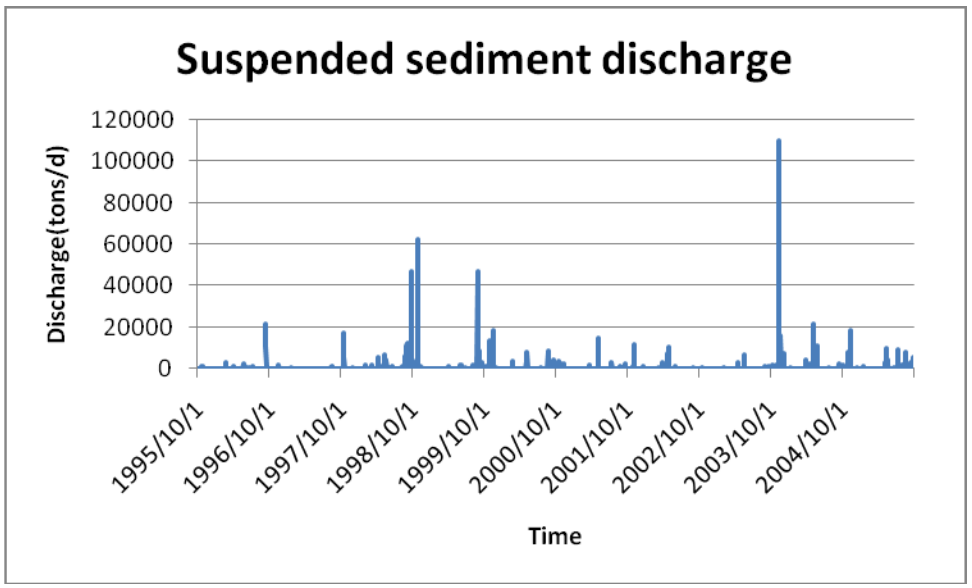
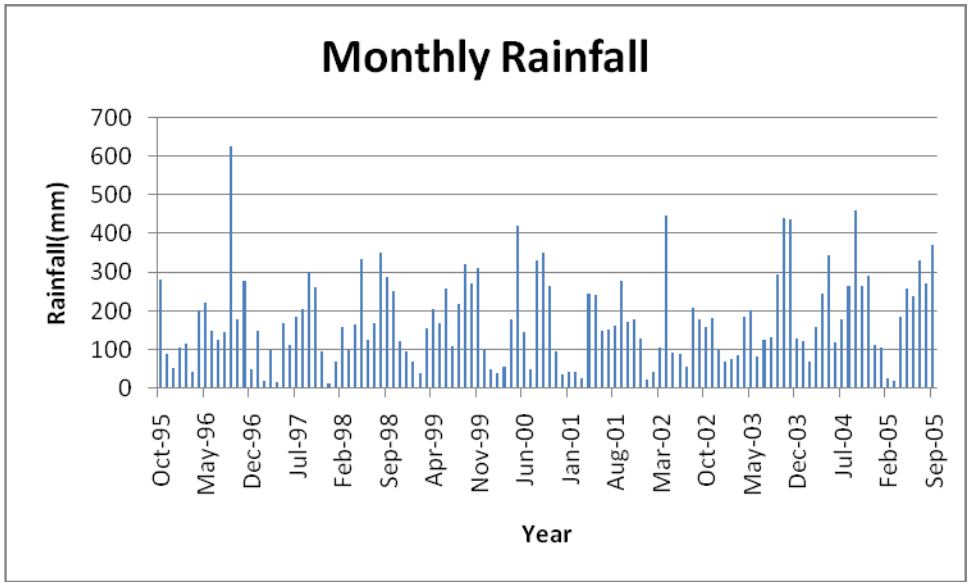


Figure 10: Daily suspended sediment discharge in Rio Mameyes

Table 9, and Figures 11, 12 show monthly rainfal and discharge data from October 1995 to September 2005 in Rio Caonillas [21].



Figuer 11: Monthly rainfall in Rio Caonillas

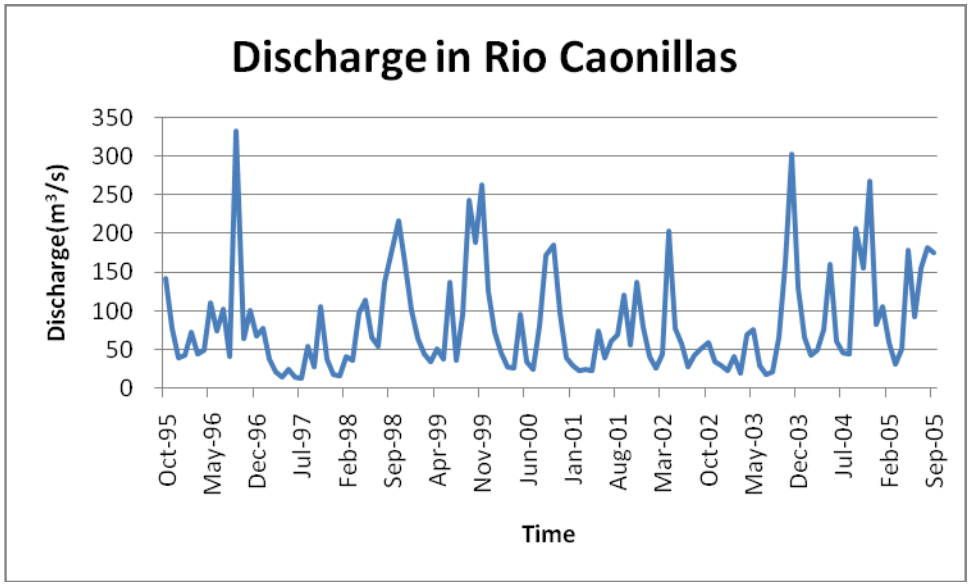


Figure 12: Monthly discharge in Rio Caonillas

1.2.2 Remote sensing data

The Landsat program provides continual global coverage with multispectral data in the visible and near infrared [22]. The Landsat Thematic Mapper (TM) sensor was carried onboard Landsats 4 and 5 from July 1982 to May 2012. Landsat 4-5 TM image data files consist of seven spectral bands. The pixel size is 30-meters (prior to February 25, 2010: thermal Band 6=60 meters). Map projection is UTM – WGS 84. Landsat 5 was launched in 1984 and continued to acquire imagery, in 7 spectral bands until November 2011 [23].

In this project, I used Standard Terrain Correction (Level 1T) TM5 images, which provide systematic radiometric and geometric accuracy by incorporating ground control points while employing a Digital Elevation Model (DEM) for topographic accuracy [24].

1.2.3 Methods

As we have seen, one of the main difficulties in the development of hydrologic and sediment transport models in the Madre de Dios basin rests in the lack of systematic field data for their calibration. To overcome these limitations, a recently proposed methodology [8, 9, 25] is used that estimates the probability distribution of flow rates in a river reach on the basis of a few physical parameters, such as soil and storm properties. To assign the values of the parameters I have used a mixture of available in situ observations and remotely sensed data.

First, I used Landsat TM5 images to obtain landuse classification maps and landuse changes in the Madre de Dios catchment. I used three classification procedures in a comparative manner: K-Means, the Iterative Self-Organizing Data Analysis Technique (ISODATA), and Spectral Angle Mapper (SAM). Second, using parameters assigned on the basis of classifications I applied the stochastic hydrologic model introduced by Botter et al. [8, 9, 25]. Finally, using a derived distribution approach I inferred the probability distribution of sediment flow, which is based on rainfall and suspended sediment concentration data.

1.2.4 Remote Sensing Data Processing Procedures

1.2.4.1 Radiometric, Geometric and Atmospheric corrections

Radiometric correction is applied to retrieve the value of the measured radiance from the binary representation of the remotely sensed data. This is done using a calibration curve given by the data provider [26]. Geometric correction is a procedure to remove the distortions caused by imperfect acquisition procedures and viewing geometry. Because TM5 images are provided with geometric and radiometric corrections, the remote sensing analyses started with atmospheric correction.

Radiation interacts with the atmosphere along the path between the ground and the sensor. Atmospheric correction is the process of removing this effect from remotely sensed data [27]. There are several algorithms to perform an atmospheric correction. The

simplest and commonly used method is dark body subtraction. This method can be used when two conditions are met: (i) The atmosphere is relatively homogeneous over the image; (ii) A black (zero reflectance) object exists within the image, often a deep water body. In the present case, because the suspended sediment concentration in the river is very large, its water appears quite bright and it is problematic to find a good approximation to a dark body. Therefore, this method is not suitable in this region [28].

I use here the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm. FLAASH corrects wavelengths in the visible through near-infrared and shortwave infrared regions, up to 3 μm. Unlike other atmospheric correction programs that interpolate radiation transfer properties from a pre-calculated database of modeling results, FLAASH incorporates the MODTRAN4 radiation transfer code [29]. FLAASH also corrects for the adjacency effect (pixel mixing due to scattering of surface-reflected radiance) and compensates for the presence of clouds [30, 31]. FLAASH is based on the following expression for the at sensor radiance:

$$L = \left(\frac{A\rho}{1-\rho_e S}\right) + \left(\frac{B\rho_e}{1-\rho_e S}\right) + L_a$$

where:

ρ is the pixel surface reflectance;

ρ_e is an average surface reflectance for the pixel and a surrounding region;

S is the spherical albedo of the atmosphere;

L_a is the radiance back scattered by the atmosphere;

A and B are coefficients that depend on atmospheric and geometric conditions but not on the surface.

Each of these variables depends on the spectral channel; the wavelength index has been omitted for simplicity.

The distinction between ρ and ρ_e accounts for the adjacency effect (spatial mixing of radiance among nearby pixels) caused by atmospheric scattering.

The values of A, B, S and La are determined from MODTRAN4 calculations that use the viewing and solar angles and the mean surface elevation of the measurement, and assume one of a set of standard model atmospheres, aerosol types, and visible range [32].

FLAASH requires the specification of scene and sensor information, including the scene center location (lat/lon), the average ground elevation of the scene, the sensor type, the sensor altitude, and the flight date and time. These data allow the computation of the sun position and the path of sunlight through the atmosphere to the ground and back to the sensor.

The next step is to select atmospheric model settings. Available model atmospheres are described in Tables 9 and 10.

In Madre de Dios, temperature is around 26 °C, and latitude is 12°36'0"S, so I selected Tropical model. The next step is to selecting an Aerosol Model:

The model choices are:

1) Rural: Represents aerosols in areas not strongly affected by urban or industrial sources. The particle sizes are a blend of two distributions, one large and one small.

2) Urban: A mixture of 80% rural aerosol with 20% soot-like aerosols, appropriate for high-density urban/industrial areas.

3) Maritime: Represents the boundary layer over oceans, or continents under a prevailing wind from the ocean. It is composed oftwo components, one from sea spray and another from rural continental aerosol (that omits the largest particles).

4) Tropospheric: Applies to calm, clear (visibility greater than 40 km) conditions over land and consists of the small-particle component of the rural model.

Due to the dominant land uses in the MDD the Aerosol Model could be rural, urban and tropospheric. Hence, I performed the atmospheric correction on the seven images covering the MDD separately. I selected the urban atmospheric model in Puerto Maldonado, and used the rural model elsewhere.

The last step is to estimate a visibility value. For Landsat, FLAASH will estimate visibility from the data itself, but an initial estimate is required. This is e.g. useful to account for heterogeneous atmospheric conditions. For example, one of the scenes used to cover the area has a thick haze (almost 20 percent cloud cover). Therefore, the initial

visibility value is set to less than 15km in this scenes. The rest of the scenes have clear weather condition, and the initial visibility value is set between 40 and 100 km.

Figure 13: Before and after atmospheric correction

Figure 13 shows details of the scenes before and after atmospheric correction.

1.2.4.2 Classification

Classification procedures are algorithms that allow to assign each pixel in remotely sensed scene to one of a set of land cover classes of interest [33]. For the particular objectives of the present thesis the classes of interest have been identified as water, forest, urban area, and agricultural fields.

Classification algorithms include unsupervised and supervised classifiers. Unsupervised classification is a technique that searches for natural groupings, or clusters, of pixels based on the similarities of their spectral signatures.

K-means and ISODATA are two commonly used unsupervised classification methods.

"K-Means unsupervised classification calculates initial class means evenly distributed in the data space, then iteratively clusters the pixels into the nearest class using a minimum distance technique. Each iteration recalculates class means and reclassifies pixels with respect to the new means. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached" [34, 35, 36].

Therefore, there are four steps in the K-Means procedure:

Step 1: Arbitrarily choose from the given sample a set of K initial cluster centers.

Step 2: According to the distances between a sample and the mean spectral signatures of all the clusters, this sample is classified as belonging to the nearest cluster.

Step 3: Calculate the new means to be the centroids of the observations in the new clusters.

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

Since the arithmetic mean is a least-squares estimator, this also minimizes the within-cluster sum of squares (WCSS) objective [37].

Step 4: Iteratively cluster the pixels into the nearest class.

This method is simple. However, the number of clusters K needs to be assigned based on some prior knowledge and the K-means outcome is sensitive to the initial starting values chosen.

The ISODATA (Iterative Self-Organizing Data Analysis Technique) approach is a modification of the k-means clustering algorithm [37]. Like K-means it randomly chooses initial 'seeds' to define information classes, then iteratively clusters the remaining pixels using minimum distance techniques. Each iteration recomputes the means and reclassifies pixels with respect to the new means [39].

Because the number of cluster center is unknown, ISODATA may have better result than K-means. Therefore, I decided to use ISODATA to get another classification map, which can compare with the result of K-means.

The ISODATA classifier requires the specification of 4 pre-specified parameters:

K_0= desired number of clusters;

n_{min}= minimum number of samples in each cluster (for discarding clusters);

σ^2_{max}= maximum variance (for splitting clusters);

d_{min}= minimum pairwise distance (for merging clusters).

The steps of the ISODATA algorithm are as follows:

1. Choose randomly $K=K_0$ initial mean vectors from the data set.
2. Assign each data point x to the cluster with closest mean:

$$x \in \omega_i \; if \; d(x,m_i) = \min\{d(x,m_1), \ldots, d(x,m_k)\}$$

3. Discard clusters containing too few members, i.e., if $n_j < n_{min}$, then discard ω_j and reassign its members to other clusters K=K-1.

4. For each cluster ω_j (j=1,…, K), update the mean vector

$$m_j = \frac{1}{n_j} \sum_{x_j \in \omega_j} x$$

and the covariance matrix:

$$\sum_j \frac{1}{n_j} \sum_{x_j \in \omega_j} (x - m_j)(x - m_j)^T$$

The diagonal elements are the variances along the N dimensions.

5. If $K \leqslant K_0/2$ (too few clusters), go to Steps 6 for splitting;

else if $K > 2K_0$ (too many clusters), go to Steps 7 for merging;

else go to Step 8.

6. (Split) For each cluster ω_j (j=1,…, K), find the greatest covariance.

If $\sigma^2_m > \sigma^2_{max}$ and $n_j > 2n_{min}$, then split m_j into two new cluster centers:

$$m_j^+ = m_j + \sigma_m$$

$$m_j^- = m_j - \sigma_m$$

Set K=K+1.

Go to Step 8.

7. (Merge) Compute the K(K-1)/2 pairwise Bhattacharyya distances between every two cluster mean vectors:

$$d_B(\omega_i,\omega_i)=\frac{1}{4}(m_i-m_j)^T\left[\frac{\Sigma_i+\Sigma_j}{2}\right]^{-1}(m_i-m_j)+log\left[\frac{\left|\frac{\Sigma_i+\Sigma_j}{2}\right|}{(|\Sigma_i||\Sigma_j|^{1/2})}\right],(1\le i,j\le K),i>j$$

For each of the distances satisfying $d_B(\omega_i,\omega_j)<d_{min}$, merge of the corresponding clusters to form a new one:

$$m_i=\frac{1}{n_i+n_j}[n_im_i+n_jm_j]$$

Delete m_j, set K=K-1.

8. Terminate if maximum number of iterations is reached. Otherwise go to Step 2.

As the number of clusters K can be dynamically adjusted in the process, the ISODATA algorithm is more flexible than the K-means algorithm. However, many more parameters, as listed previously, have to be assigned, usually empirically [40].

Supervised classification is a general technique that uses information derived from a few areas of known identity to classify unknown pixels in the remaining image. Some classes will be defined as training sets. From those training areas we learn about the reflection pattern of each particular class [41].

SAM [42, 43] assumes that the data have been reduced to apparent reflectance (true reflectance multiplied by some unknown gain factor controlled by topography and

shadows). The algorithm determines the similarity between two spectra by calculating the "spectral angle" between them, treating them as vectors in a space with dimensionality equal to the number of bands (nb). A simplified explanation of this can be given by considering a reference spectrum and an unknown spectrum from two-band data. The two different materials will be represented in the 2-D scatter plot by a point for each given illumination, or as a line (vector) for all possible illuminations.

The SAM algorithm generalizes this geometric interpretation to nb -dimensional space. SAM determines the similarity of an unknown spectrum t to a reference spectrum r, by applying the following equation [32]:

$$\alpha = \cos^{-1}\left(\frac{\vec{t}\vec{r}}{\|\vec{t}\|\|\vec{r}\|}\right)$$

which also can be written as:

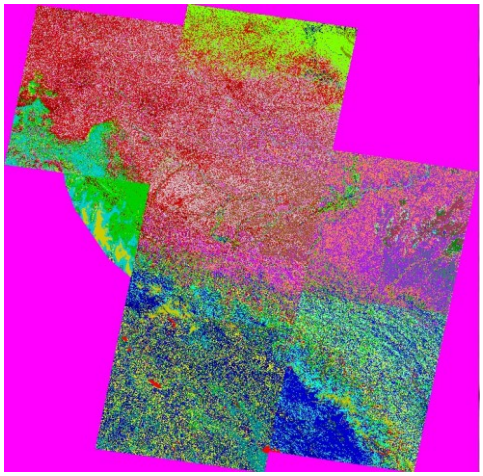
$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{nb} t_i r_i}{\left(\sum_{i=1}^{nb} {t_i}^2\right)^{1/2}\left(\sum_{i=1}^{nb} {r_i}^2\right)^{1/2}}\right)$$

where nb equals the number of bands in the image.

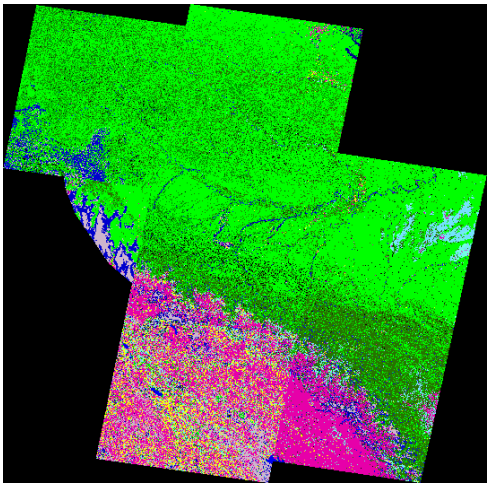
Based on this principle, I selected a few hundred training sites to cover all the classes of interest.

2. Classification Results

Figure 14a shows the K-Means classification result, which is obviously not satisfactory. For example, the water classes can hardly be separated from other landcover types.



(a)



(b)

Figure 14: a) K-means classification; b) ISODATA classification.

ISODATA classfication results (Figure 14b) are significantly better than those obtained from K-means. ISODATA successfully separates river pixels from other landcover type. However, the whole Madre de Dios region is covered using seven distinct scenes, acquired at different times. The ISODATA classifier, applied separately to each scene, produces different class distributions, e.g. due to changing incoming radiation and atmospheric states. As a result, neighboring pixels in different scenes are often attributed to different classes, such that the boundaries between the scenes are very evident.

Figure 15: One of the training sites

The SAM classifier yields very realistic results (Figure 16), in which river locations are clearly distinguishable and boundaries between scenes are barely noticeable.

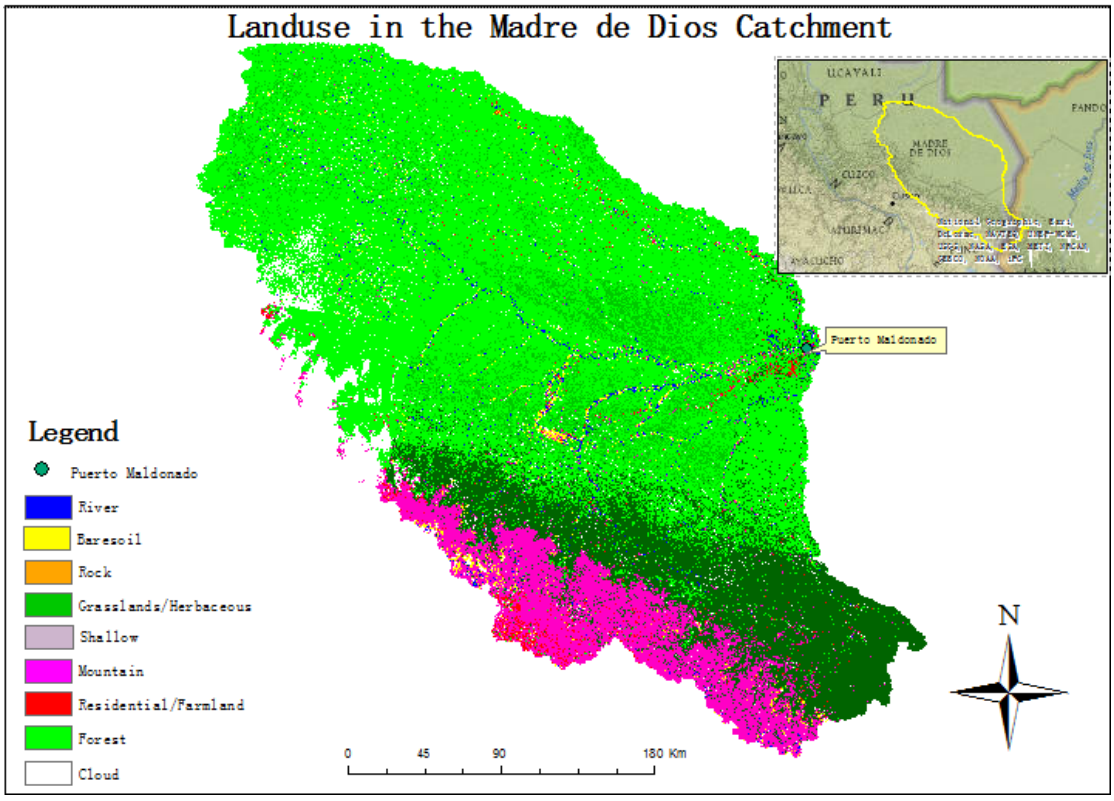
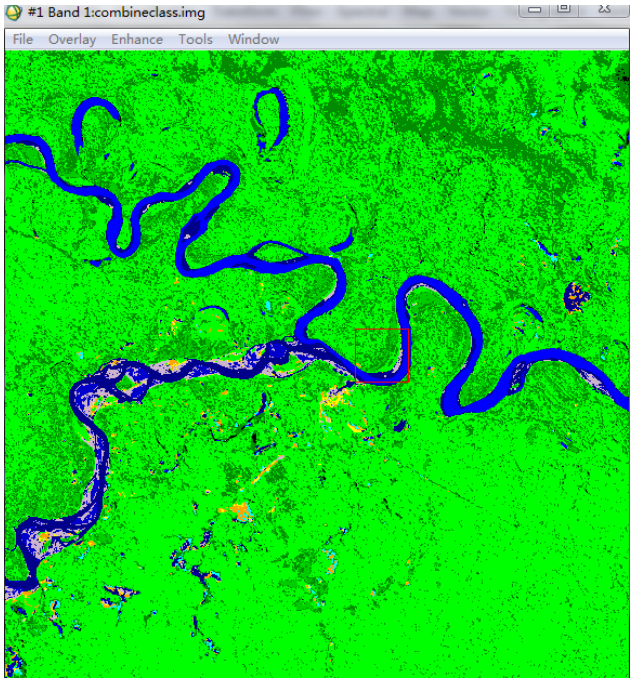


Figure 16: SAM classification by ENVI.

To illustrate the capabilities of remote sensing in characterizing surface features, Figure 17 shows a zoomed-in portion of the MDD, where water of two "types" can be identified: reaches in which suspended sediment concentration exhibit a sharp spatial contrast. It is indeed common to observe such a contrast in many sections along the MDD river (Figure 17a). This classification is based on photographs in this area, which should be verified by observed sediment concentration data in the future.

(a)

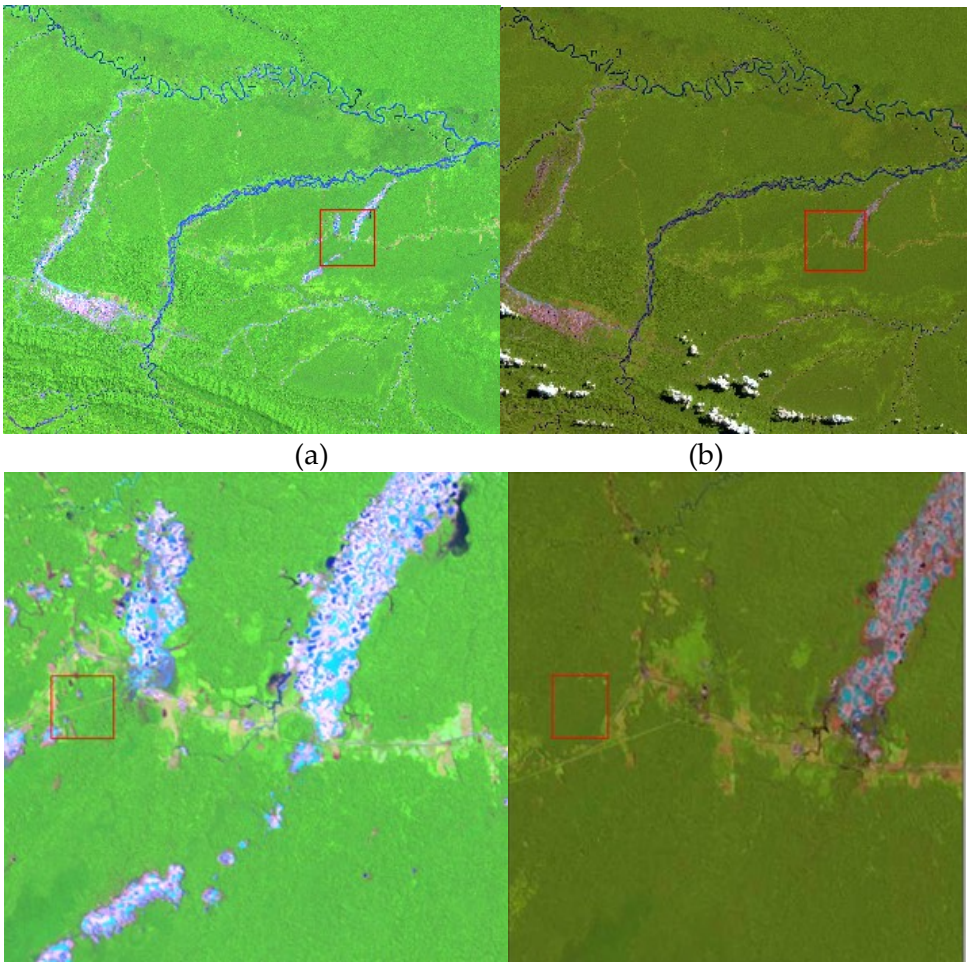


(b)

Figure 17: Detail of the SAM classification. The dark blue river rich has a much higher suspended sediment concentration than the remaining river section, as illustrated in the inset (by Chauca, source: Google Earth).

Remote sensing can also be used to document the time evolution of gold mining areas [44, 45]. Figure 18 shows how the forest clearings associated with gold mining

have dramatically increased in area from 2009 to 2011 in the southern portion of the Madre de Dios region.



(a) (b)

(c) (d)

Figure 18: Expanding gold mining areas. a) and c) were acquired in 2011, b) and d) in 2009

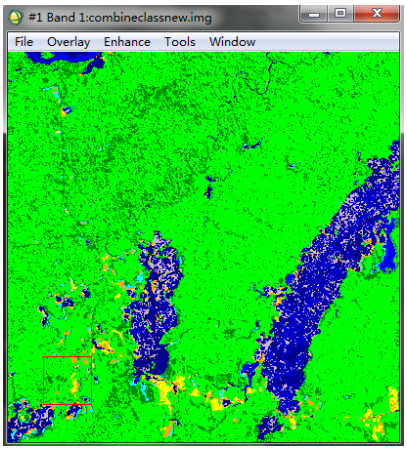


Figure 19: SAM classification image show the new gold mines in 2011.

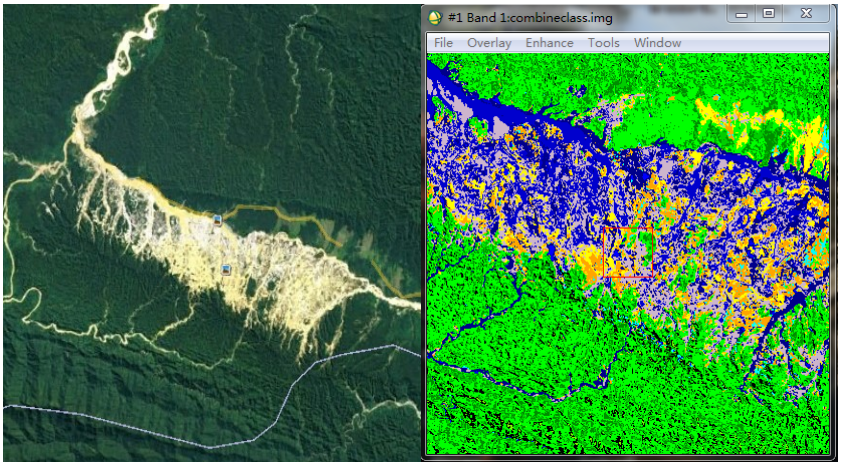


Figure 20: The largest gold mine in Madre de Dios region in 2011.

Figure 20 shows there are large areas with standing water (pink area) associated with forest clearings in the gold mining areas. Mined area provides high contrast with the background forest. Although the mining areas may be confused with river features, I selected training sites to alleviate this problem. Classes are separated by visual interpretation based on the image bands, tasseled-cap indices and photographs. Mining area included a range of subclasses from water bodies to exposed soil surfaces [44].

3. Streamflow and sediment flow probability distributions

3.1 Theoretical Background

In order to quantify the impact of gold mining on mercury transports, it is crucial to estimate the frequency distribution of streamflow and sediments flows in the MDD river. Botter [8, 9] has proposed physically-based relationships to link stochastic streamflow fluctuations and rainfall and soil moisture dynamics. The steady state probability distribution function (pdf) of the subsurface contribution to streamflows can be explained, under some simplifying assumptions, by rainfall, soil, vegetation and geomorphological features. The approach introduces the following parameters: root zone depth (i.e., the depth of the active soil layer), Zr [L]; porosity, n; and maximum evapotranspiration rate, ET [L/T]. Losses due to evapotranspiration are assumed to increase linearly from 0 at the wilting point, s_w, up to the maximum evapotranspiration rate, ET, at a suitable soil moisture threshold, s_1 [46, 47].

The temporal evolution of spatially‐averaged relative soil moisture in the root zone, s(t), is thus seen as the result of the following three processes [47, 48]: (1) stochastic instantaneous increments due to infiltration from rainfall; (2) linear losses due to evapotranspiration; (3) deep percolation producing effective rainfall and subsurface contributions to streamflow (above the threshold s_1).

Stochastic instantaneous increments due to infiltration from rainfall are modeled, at daily timescales, as a Poisson process with average frequency λ_P [T^{-1}] (and where daily

rainfall depths are assumed to be exponentially distributed with parameter γ_P [L^{-1}]) [9, 25, 47].

The related average frequency of runoff events, λ, is expressed in terms of the underlying soil, vegetation and rainfall properties as [49]:

$$\lambda = \eta \frac{\exp(-\gamma_s)\gamma_s^{\frac{\lambda_p}{\eta}}}{\Gamma(\frac{\lambda_p}{\eta}, \gamma_s)} \quad (1)$$

where $\Gamma(a, b)$ is the lower incomplete gamma function. The ratio between the soil storage capacity and the mean rainfall depth is γ_s:

$$\gamma_s = \gamma_p n Z_r (s_1 - s_w)$$

The normalized maximum evapotranspiration rate is finally defined as:

$$\eta = ET/(nZ_r s_1 - s_w)$$

3.2 Linear model

If discharge is assumed to be a linear function of the volume of water stored within the basin [8, 9, 24], the temporal evolution of streamflows, Q, can be described as:

$$\frac{dQ(t)}{dt} = -kQ + \xi_t(\lambda; \gamma) \quad (2)$$

where streamflow pulses are due to percolation events (described by the term ξ_t) and follow an exponential distribution of parameter $\gamma = \gamma_P/(kA)$ (A being catchment area [9]).

The analytical streamflow pdf resulting from the linear model represented by eq. (2) can be expressed as:

$$p(Q) = \frac{\gamma_Q'^{\lambda/k}}{\Gamma(\lambda/k)} Q^{(\frac{\lambda}{k}-1)} \exp(-\gamma_Q' Q) \quad (3)$$

With

$$\gamma_Q' = \frac{\gamma_Q}{1 - e^{-r}}$$

γ_Q is defined as:

$$\gamma_Q = \frac{\gamma_p}{nkAZ_r}$$

$$r = \gamma_p(1 - s_1)$$

According to the linear streamflow pdf function, Basso and Botter [52] derived the sediment concentration pdf function:

$$p(Cs) = \left(\frac{1}{a * (\alpha * k)^d} * Cs\right)^{\frac{\lambda}{k*d}} e^{-a*(\alpha*k)^d * Cs^{\frac{1}{d}}} \frac{1}{a * (\alpha * k)^d d * gamma\left(\frac{\lambda}{k}\right)}$$

where Cs is sediment concentration, parameters a and d express the relationship between sediment concentration (Cs) and stream discharge (Q) [50]:

$$Cs = aQ^d \tag{4}$$

3.3 Nonlinear model

In the more realistic case in which the relationship between stored volume and discharge is nonlinear [50], the analytical streamflow pdf can be found by solving:

$$\frac{dQ(t)}{dt} = -kQ(t)^\alpha + \xi_t \tag{5}$$

where ξ_t represents stochastic noise (the jumps of Q in correspondence of the fraction of rainfall events producing streamflow); α and k [$L^{1-\alpha}T^{\alpha-2}$] are constants.

The general expression of the steady state probability distribution function of streamflow from (5) is the following [9]:

$$p(Q) = N\left\{\frac{1}{Q^\alpha} exp\left[-\frac{\gamma_w}{k(2-\alpha)}Q^{2-\alpha} + -\frac{\lambda}{k(1-\alpha)}Q^{1-\alpha}\right] + \frac{k}{\lambda}\delta(Q)H[1-\alpha]\right\}$$

where N is a normalizing constant, H is the Heaviside unit step function, $\gamma_W=\gamma_P/A[L^{-1}]$ represents the inverse of the mean storage increment due to incoming rainfall events, and $\lambda[T^{-1}]$ represents the average frequency of effective rainfall (i.e., runoff) events [25].

According to the nonlinear streamflow pdf function, Basso and Botter [52] derived the sediment concentration pdf function:

$$p(Cs) = N * \frac{1}{d} * \frac{1}{a}^{\frac{1-\alpha}{d}} Cs^{\frac{1-d-\alpha}{d}} e^{\left(\frac{-\gamma_w * \left(\frac{1}{a}\right)^{\frac{2-\alpha}{d}}}{k*(2-\alpha)}\right) * \left(Cs^{\frac{2-\alpha}{d}}\right) + \left(\lambda \frac{\frac{1}{a}^{\frac{1-\alpha}{d}}}{k*(1-\alpha)}\right) * \left(Cs^{\frac{1-\alpha}{d}}\right)}$$

where Cs is the sediment concentration, parameters a and d express the relationship between sediment concentration (Cs) and stream discharge (Q) in eq. (4).

3.4 Parameters

The parameters required to define the probability distribution function of streamflows for a river basin includes: mean observed streamflow (Q_{mean}), mean rainfall frequency (λ_p), and the inverse of the mean daily rainfall depth (γ_w).

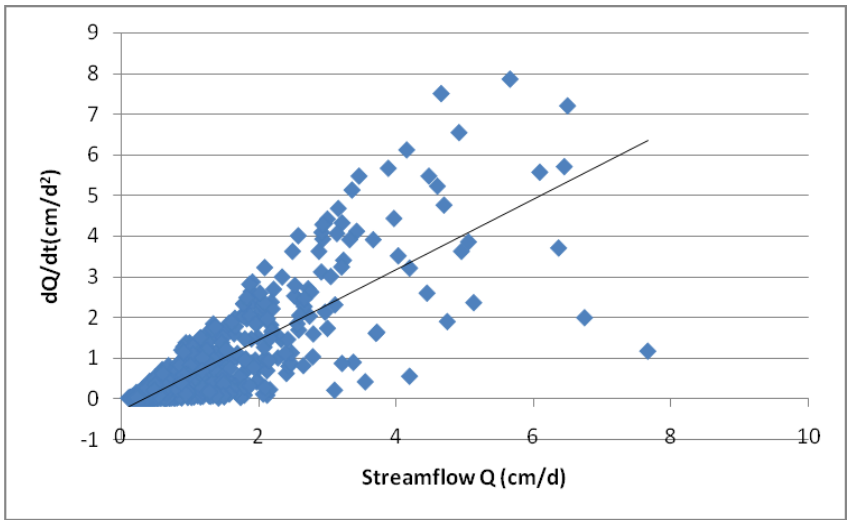
Q_{mean} is here estimated from observed daily streamflow discharge Q_{obs}. λ_p is estimated directly from streamflow data, by simply counting rainfall events in the time series available. γ_w=1/a′ is estimated simply by computing a′ =mean rainfall depth.

The average frequency of streamflow events, λ, is estimated by dividing the observed mean specific discharge, Q_{mean}, by the mean rainfall depth event, a′ [53]:

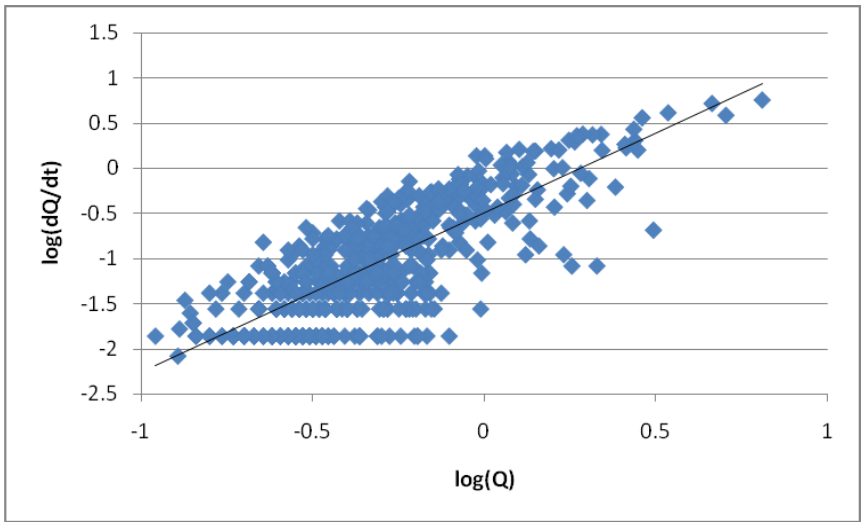
$$\lambda = Q_{mean}/a' \tag{6}$$

where a′ can be computed from rainfall records as the observed mean daily depth. Botter has verified the estimate of λ through the comparison of theoretical and observed pdfs and by comparing the estimated value of λ with the rainfall frequency λ_p. He found eq. (1) and eq. (6) has the similar results to describe the rainfall attributes [53].

With this estimation procedure direct knowledge of the evapotranspiration rate is not needed.



(a)



(b)

Figure 21: Comparing daily streamflow distribution in the Rio Mameyes. a) Recession Q plot with flow recession rates; b) Recession log-log plot with flow recession rates (individual daily data, dots) and regression line (solid line)

According to eq. (1), i.e. the linear model with α=1, k can be estimated from streamflow measurements by a linear regression. The value of k can, in fact, be found by plotting dQ/dt vs. the corresponding observed value of Q [24]. According to eq. (4), in the nonlinear model, k and α are estimated by linear least squares regression of the average log(−dQ/dt) plotted versus the corresponding average observed values of log(Q) [25].

3.5 Application to the MDD

Because we only have rainfall data in the Madre de Dios catchment, and no discharge observation, the application to the MDD requires different estimation procedures.

I use the Rational Method, which states that the discharge, Q, at the catchment outlet point can be calculated as [54]:

$$Q = C\,I\,A$$

where I is the rainfall intensity, A is the drainage basin area, and C is the runoff coefficient.

Casimiro provides an estimated runoff coefficient C= 0.82 for a neighboring catchment (15.26°S, 69.87°W), with similar characteristics to the MDD [55].

The statistical relationship between suspended sediment concentration and stream discharge is called the rating curve [50, 56] and, as before, it is here assumed to take the power law form:

$$Cs = aQ^d$$

The suspended load Qs of a river is similarly related to the discharge by the same rating coefficients [50, 56]:

$$Qs = aQ^{d+1}$$

Syvitski provided regression equations for coefficients a and d [50]:

$$log(a) = 0.55 - 1.66log\ (Q) - 0.00029R$$

$$d = 0.90 - 0.026T + 0.00020R$$

where R is basin relief, T is mean annual air temperature.

3.6 Results and discussion

3.6.1 Streamflow and sediment transport estimates in Puerto Rico

Based on the discussion above, we can obtain the relevant parameter as follows:

Table 1: Summary of the relevant parameters in station 1: inverse of the mean daily rainfall depth, γ_w; mean rainfall frequency, λ_p; mean observed streamflow, Q_{mean}.

Season	Q_{mean}(cm/d)	λ_p(d^{-1})	γ_w(cm^{-1})
Spring	0.593	0.843	0.957
Summer	0.675	0.951	0.814

Autumn	0.902	0.932	0.622
Winter	0.741	0.912	0.778
Year	0.723	0.910	0.775

Table 2: Summary of the relevant parameters in station 2: inverse of the mean daily rainfall depth, γw; mean rainfall frequency, λp; mean observed streamflow, Q_{mean}.

Season	Q_{mean}(cm/d)	λ_p(d^{-1})	γ_w(cm^{-1})
Spring	0.481	0.539	0.937
Summer	0.471	0.642	1.133
Autumn	1.202	0.733	0.844
Winter	0.419	0.516	2.085
Year	0.643	0.608	1.076

Table 3: Parameters of the stochastic Models in station 1: the related average frequency of runoff events λ, coefficients of temporal decay of Q, k and α.

Season	Method	λ(d^{-1})	α	k($cm^{1-\alpha}d^{2-\alpha}$)
Spring	Nonlinear	0.568	2.718	0.732
	Linear		1	0.344
Summer	Nonlinear	0.550	1.837	0.496
	Linear		1	0.318
Autumn	Nonlinear	0.561	2.192	0.375
	Linear		1	0.385
Winter	Nonlinear	0.577	2.210	0.341
	Linear		1	0.265
Year	Nonlinear	0.560	2.262	0.530
	Linear		1	0.326

Table 4: Parameters of the stochastic Models in station 2: the related average frequency of runoff events λ, coefficients of temporal decay of Q, k and α

Season	Method	λ(d^{-1})	α	k($cm^{1-\alpha}d^{2-\alpha}$)
Spring	Nonlinear	0.450	2.334	0.577
	Linear		1	0.224
Summer	Nonlinear	0.534	2.617	0.936
	Linear		1	0.180

Autumn	Nonlinear	1.015	2.078	0.147
	Linear		1	0.191
Winter	Nonlinear	0.874	3.190	0.834
	Linear		1	0.078
Year	Nonlinear	0.691	2.538	0.439
	Linear		1	0.173

I subsequently compared the performance of the linear and nonlinear models with streamflow and suspended flux observations in the two study basins selected in Puerto Rico.

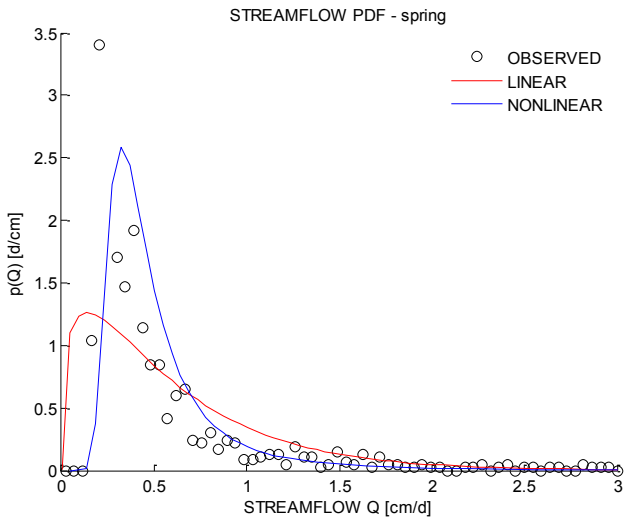


Figure 22: Streamflow pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in spring

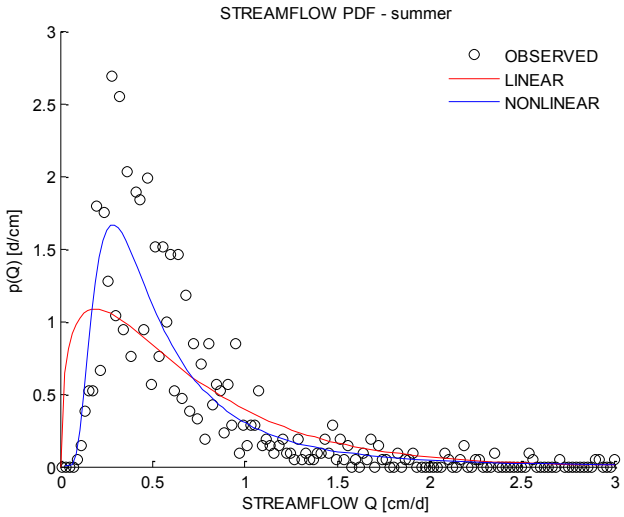


Figure 23: Streamflow pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in summer

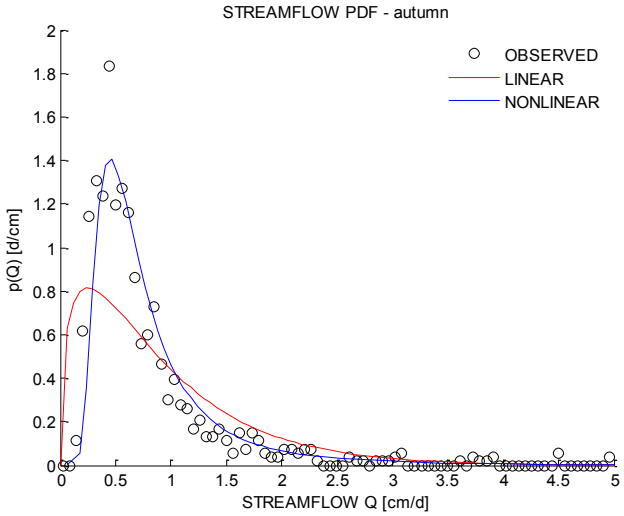


Figure 24: Streamflow pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in autumn

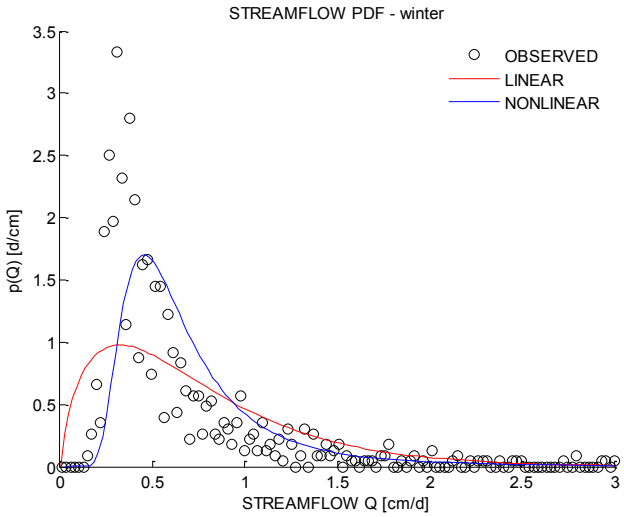


Figure 25: Streamflow pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in winter

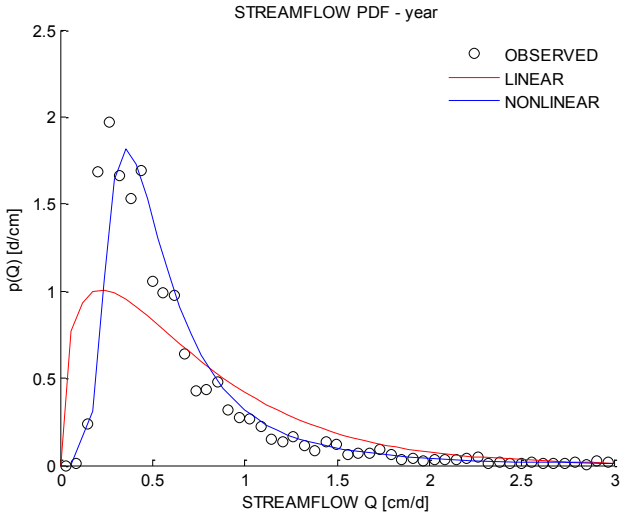


Figure 26: Streamflow pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in year.

Figure 22 to 26 show that the nonlinear model seems to better capture the behavior of the Rio Mameyes Catchment. In particular, the nonlinear model agrees

better than the linear model with observations near the peak of the distribution.

However, the value of the distribution at the mode tends to be underestimated.

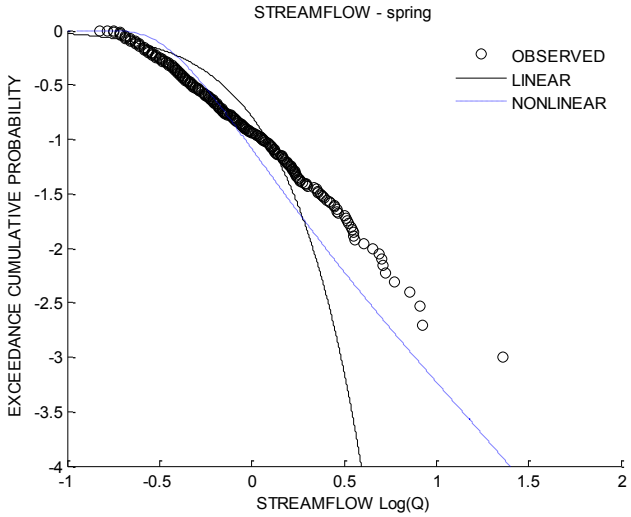


Figure 27: Streamflow cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in spring

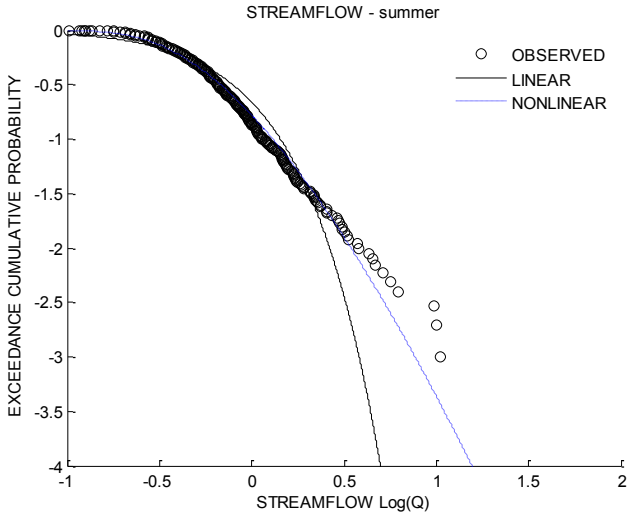


Figure 28: Streamflow cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in summer

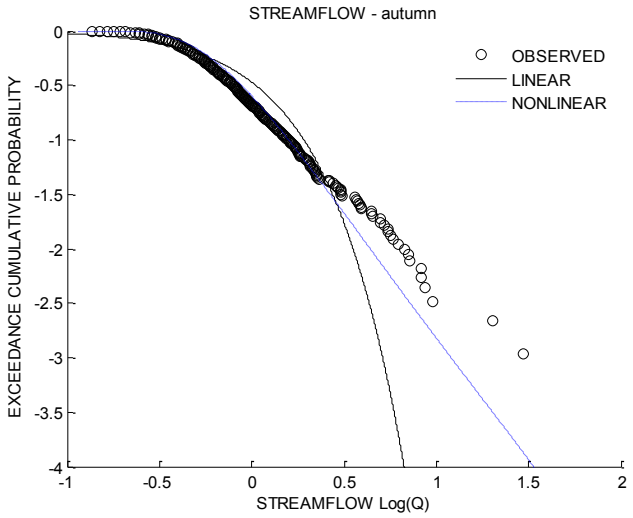


Figure 29: Streamflow cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in autumn

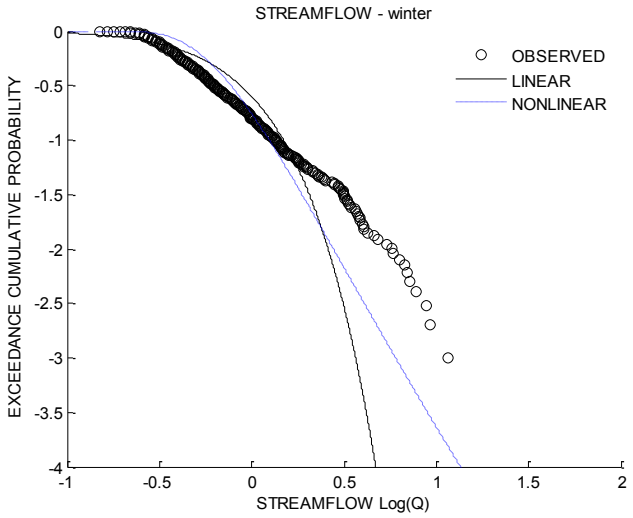


Figure 30: Streamflow cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in winter

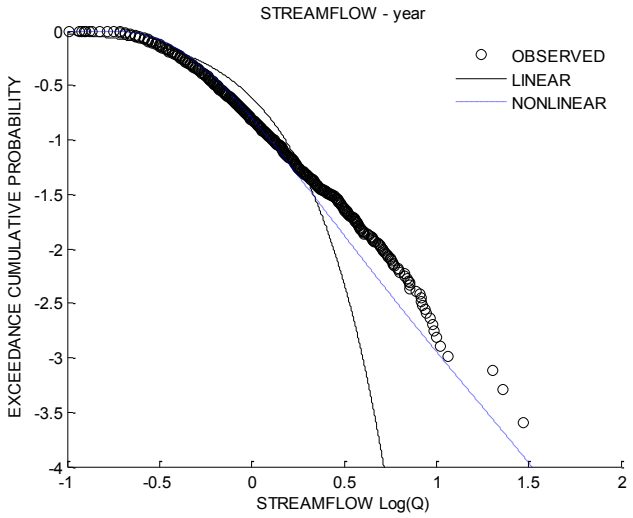


Figure 31: Streamflow cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in year

Figure 27 to 31 compare the survival probabilities obtained from the analytical stochastic formulations for Rio Mameyes and those from observations. Also this representation shows a better agreement of the pdf obtained from the nonlinear model with observations. In particular, the linear model severely overestimates the observed frequency distribution when the Log(Q) is from -0.5 to 0.5.

On the seasonal scale, the nonlinear model also displays some shortcomings. For example, the nonlinear model systematically underestimates the probability of high values of Q.

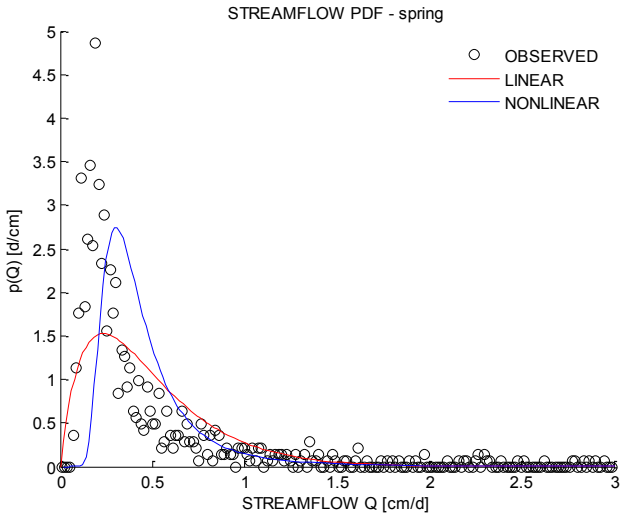


Figure 32: Streamflow pdf in Rio Caonillas: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in spring

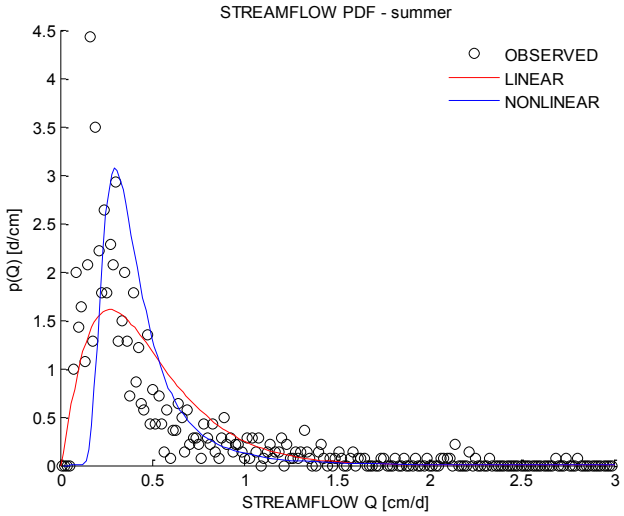


Figure 33: Streamflow pdf in Rio Caonillas: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in summer

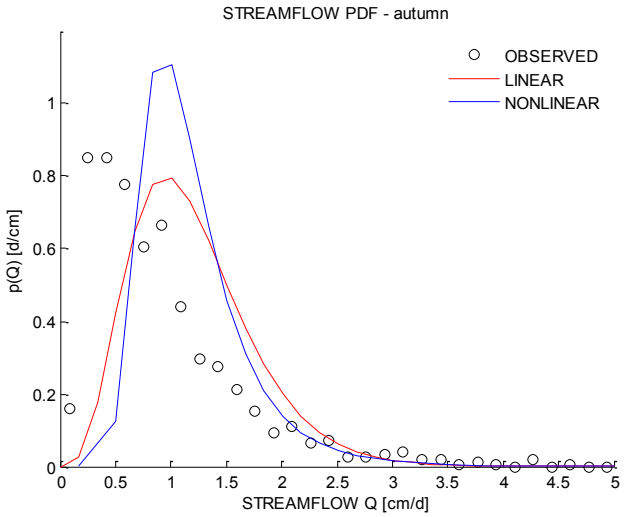


Figure 34: Streamflow pdf in Rio Caonillas: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in autumn

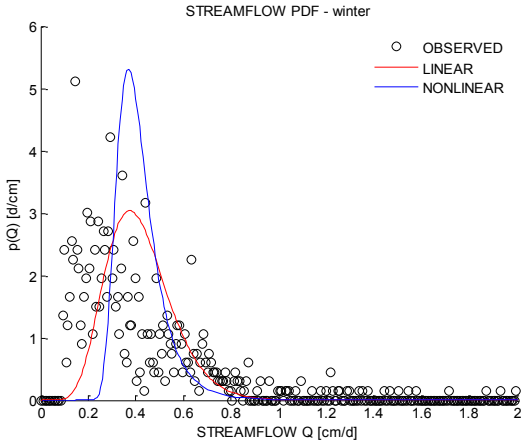


Figure 35: Streamflow pdf in Rio Caonillas: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in winter

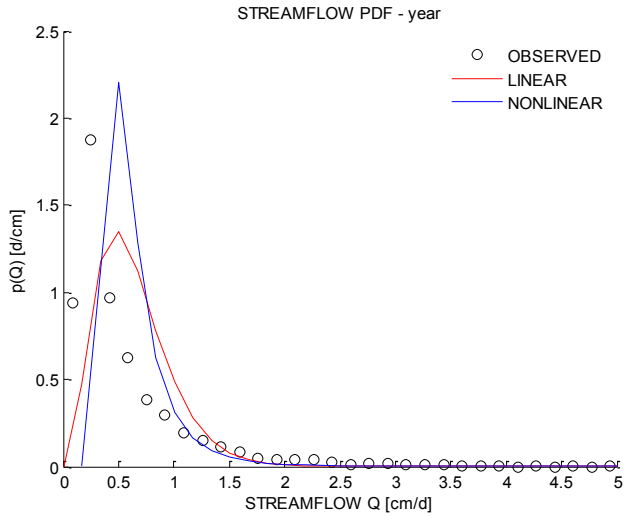


Figure 36: Streamflow pdf in Rio Caonillas: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in year.

Figures 32 to 36 show the results obtained for Rio Caonillas. Also in this case the nonlinear model better describes the observed statistics.

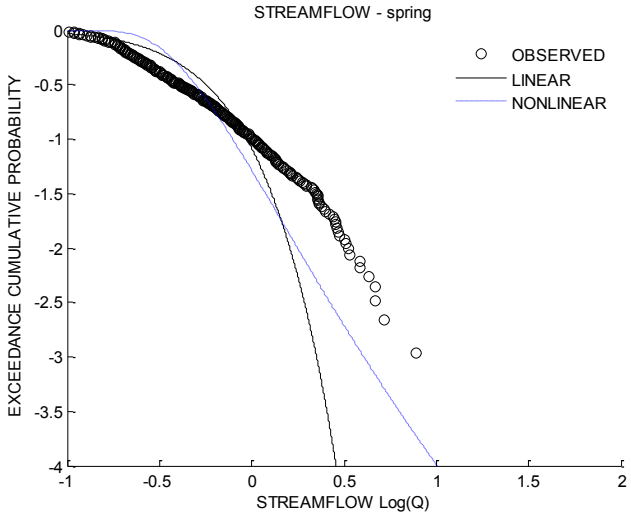


Figure 37: Streamflow cdf in Rio Caonillas: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in spring

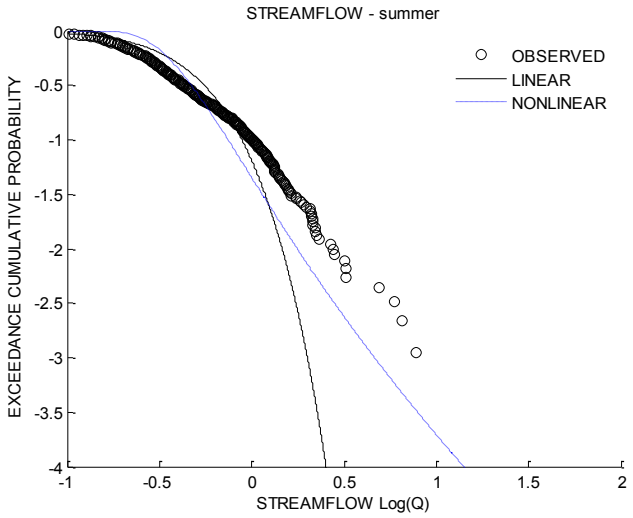


Figure 38: Streamflow cdf in Rio Caonillas: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in summer

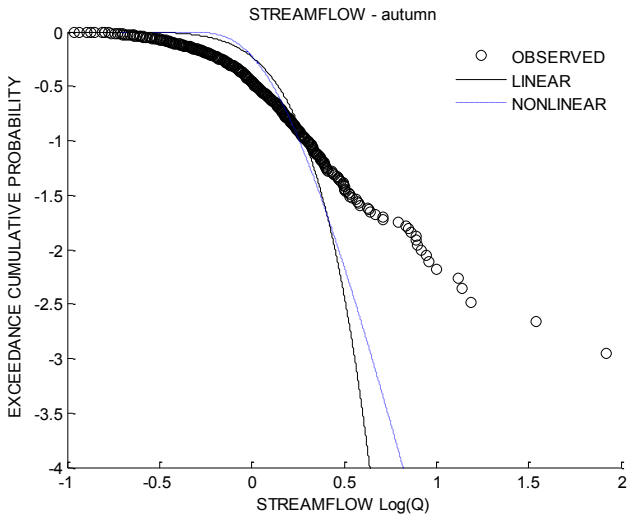


Figure 39: Streamflow cdf in Rio Caonillas: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in autumn

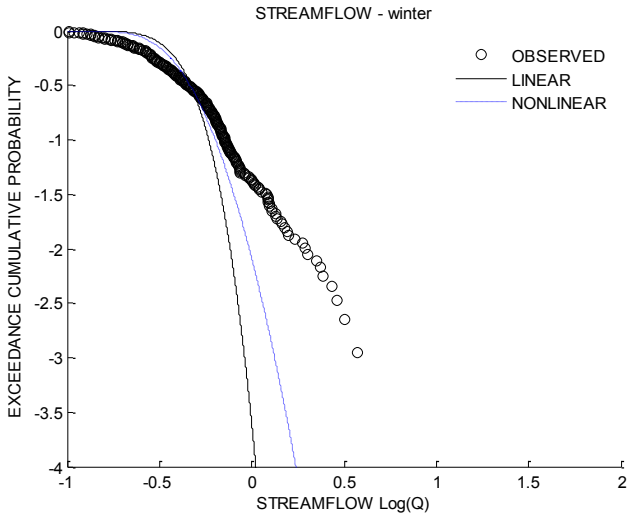


Figure 40: Streamflow cdf in Rio Caonillas: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in winter

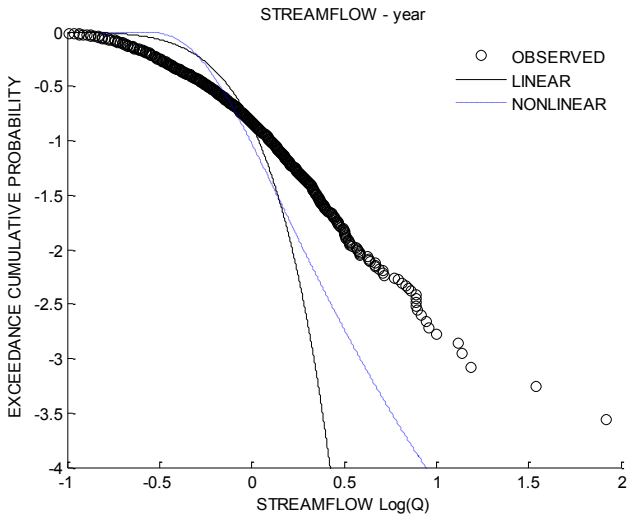


Figure 41: Streamflow cdf in Rio Caonillas: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in year

Exceedance probability distributions are in Figure 37 to 41. Also this representation shows that the nonlinear model provides better estimates of the distributions of discharge.

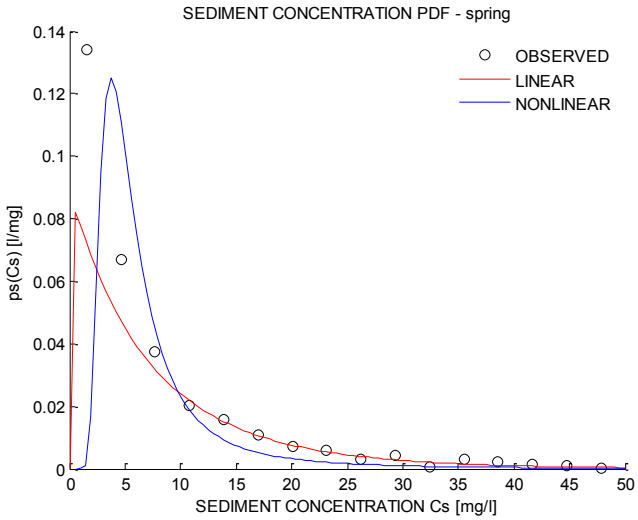


Figure 42: Sediment concentration pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in spring

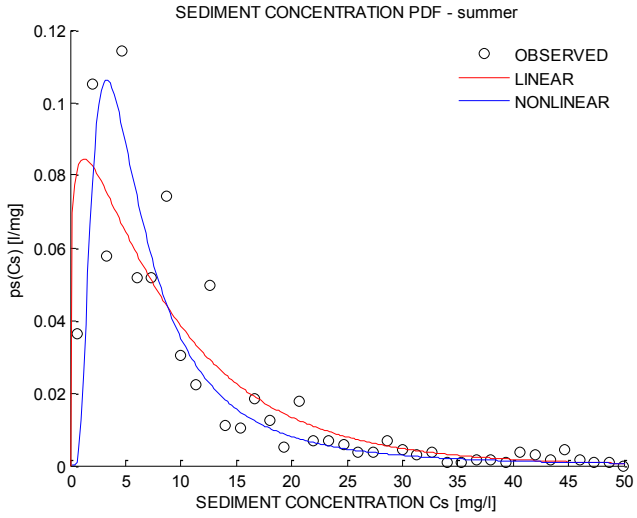


Figure 43: Sediment concentration pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in summer

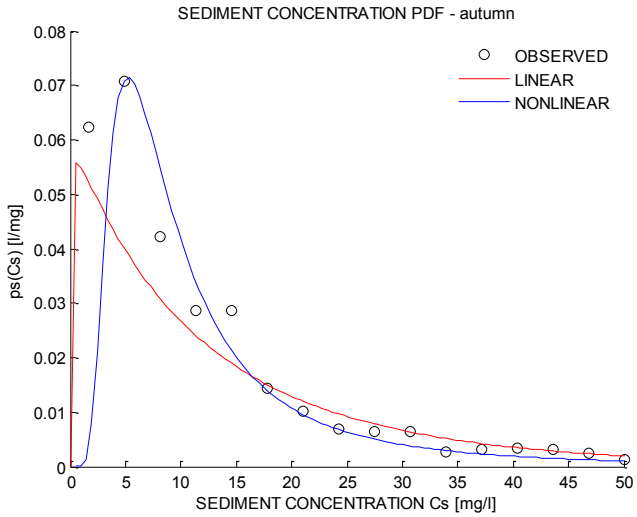


Figure 44: Sediment concentration pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in autumn

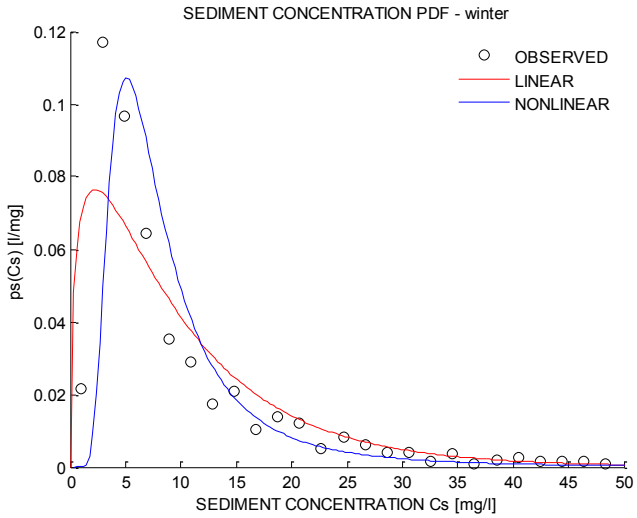


Figure 45: Sediment concentration pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in winter

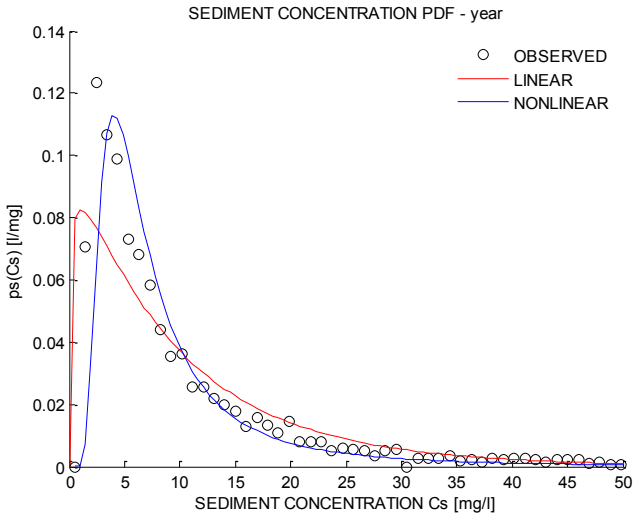


Figure 46: Sediment concentration pdf in Rio Mameyes: blue line is the nonlinear model; red line is the linear model, and observed streamflow pdf is circles in year.

Figures 42 to 46 compare the pdf for suspended sediment concentration obtained from the two stochastic models, with those from observed ones. Also for the sediment transport distribution, the nonlinear model better describes the observations.

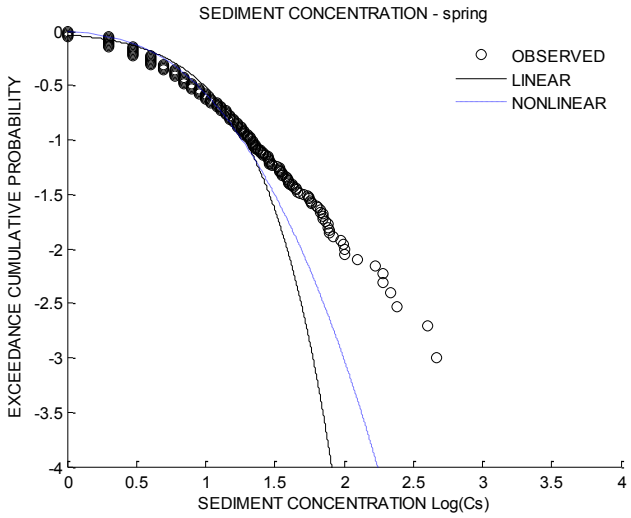


Figure 47: Sediment concentration cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in spring

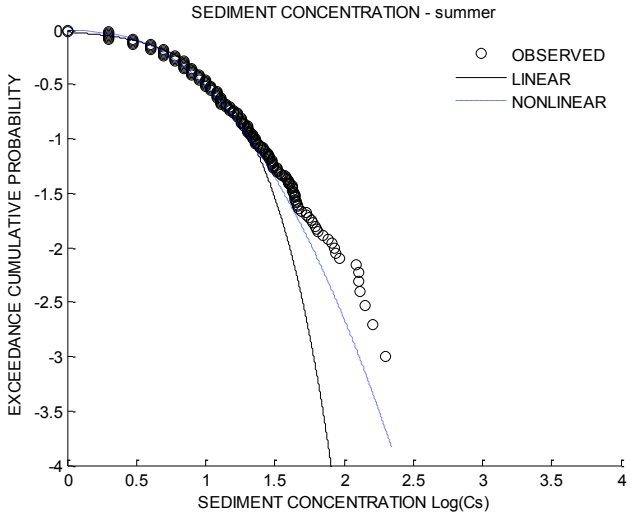


Figure 48: Sediment concentration cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in summer

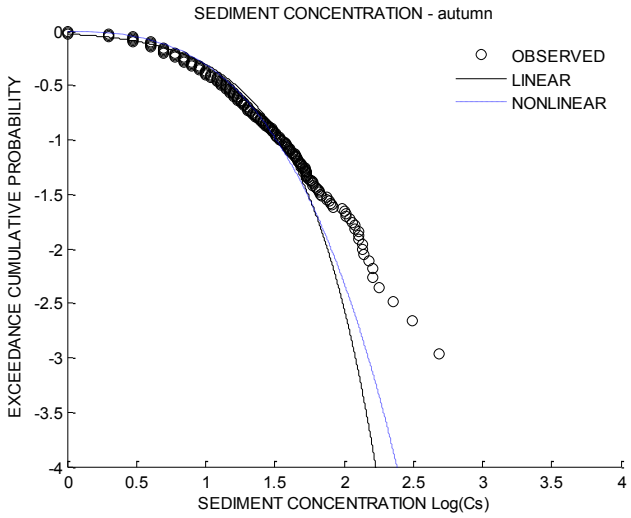


Figure 49: Sediment concentration cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in autumn

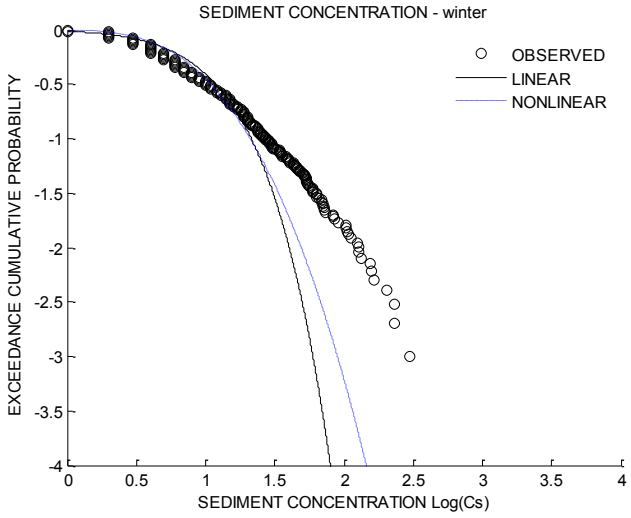


Figure 50: Sediment concentration cdf in Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in winter

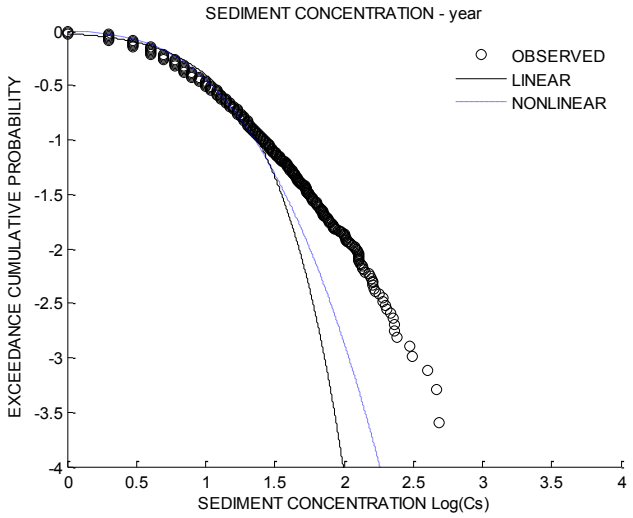
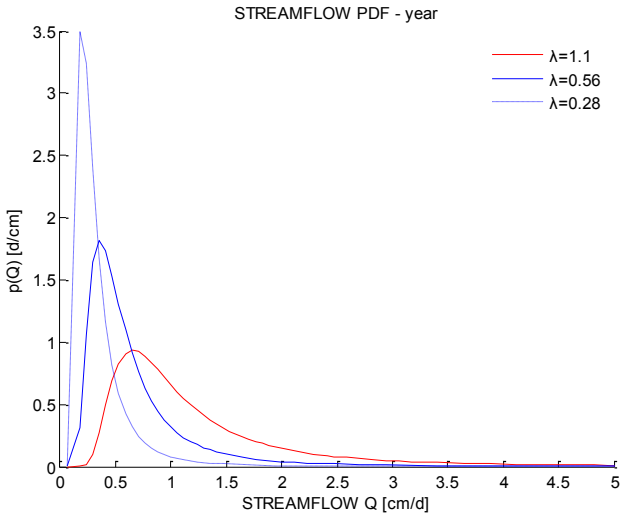


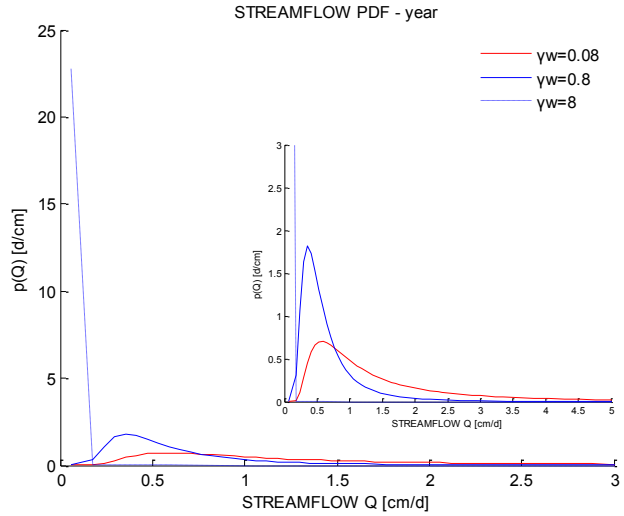
Figure 51: Sediment concentration cdf in the Rio Mameyes: dash line is nonlinear model, solid line is linear model, and observed streamflow pdf is circles in year

Figures 47 to 51 show the survival probability of sediment concentration for Rio Mameyes. Comparisons between analytical stochastic approaches and observations show a better performance of the nonlinear approach to stochastic flow and transport modeling.

On the seasonal scale the nonlinear model shows departures from observations. Figure 47 shows the linear model provides a better fit when Log(Cs) is between 1 and 1.5 in the spring season. Nevertheless, these comparisons show that predictions from the nonlinear model are closer to observed sediment concentration pdf in most cases, especially on a yearly scale. I changed λ and γ_w in a large range to see how the parameters affect the results in nonlinear model.



(a)



(b)

Figure 52: Sensitivity analysis in the case of nonlinear model. a) The unit of λ is d^{-1} b) The unit of γ_w is cm^{-1}.

Figure 52a emphasizes that decreasing λ lead to pronounced decreases of the distribution.

Figure 52b suggests that decreasing the mean rainfall rate (i.e., increasing γ_w) leads to a marked decrease of the distribution (i.e., γ_w becomes drier and narrower) [9].

Visual inspection of the figures above indicates that the model has a tendency to underestimate flow rates and, hence, sediment transport rates. In order to characterize the estimation error I evaluated the percent estimation error for both the seasonal and yearly estimates:

$$E = \frac{V_{est} - V_{obs}}{V_{obs}}$$

where V_{est} is estimated sediment volume flowing through the outlet during the period considered and V_{obs} is the observed sediment volume flowing through the outlet during the same period. Such error estimates will then be used in the ungauged MDD basin to provide a measure of uncertainty.

Table 5: The estimation error

Season	Spring	Summer	Autumn	Winter	Year
E	-0.186	-0.087	-0.049	-0.205	-0.135

3.6.2 Streamflow and sediment transport estimates in the Madre de Dios basin

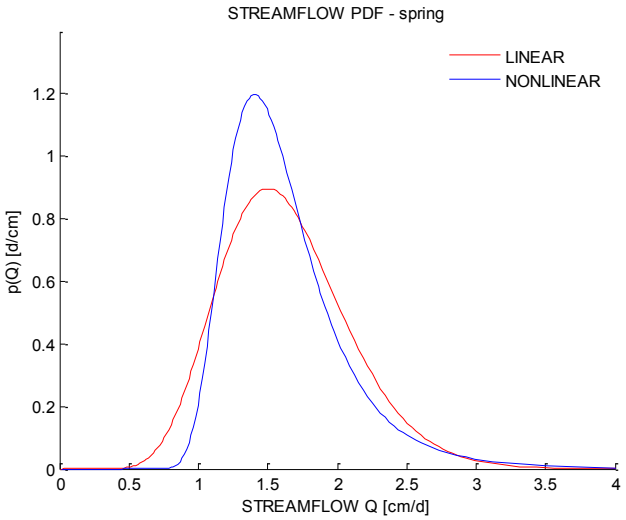


Figure 53: Streamflow pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

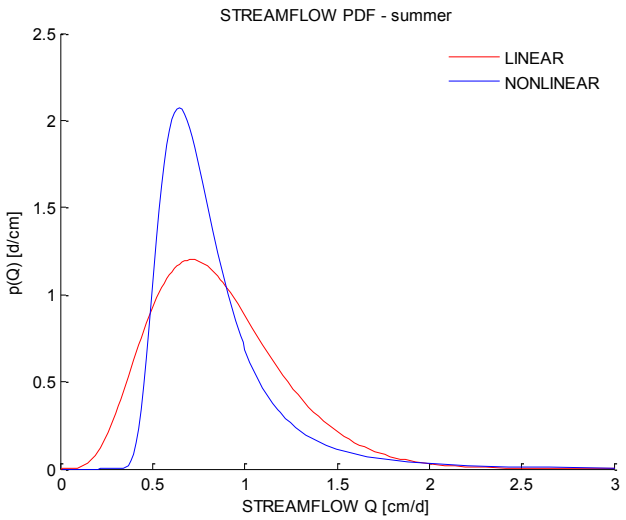


Figure 54: Streamflow pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

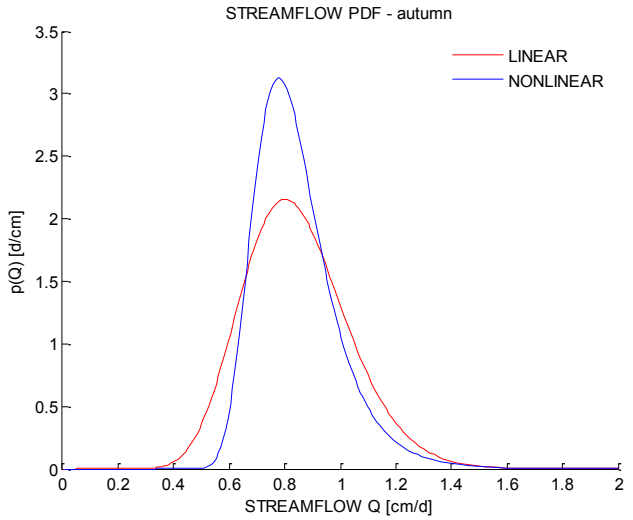


Figure 55: Streamflow pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

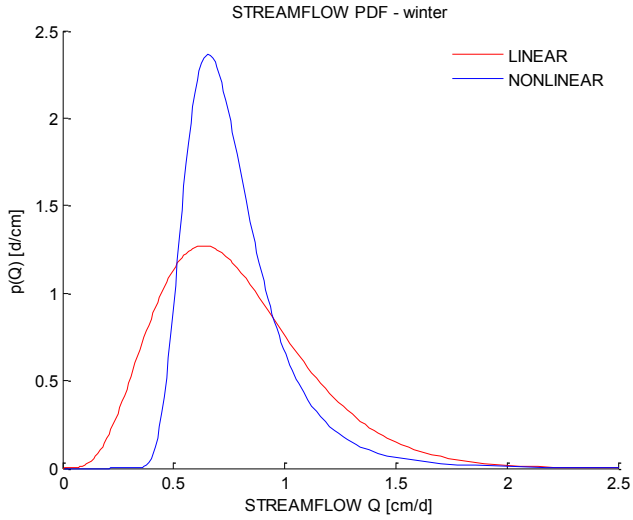


Figure 56: Streamflow pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

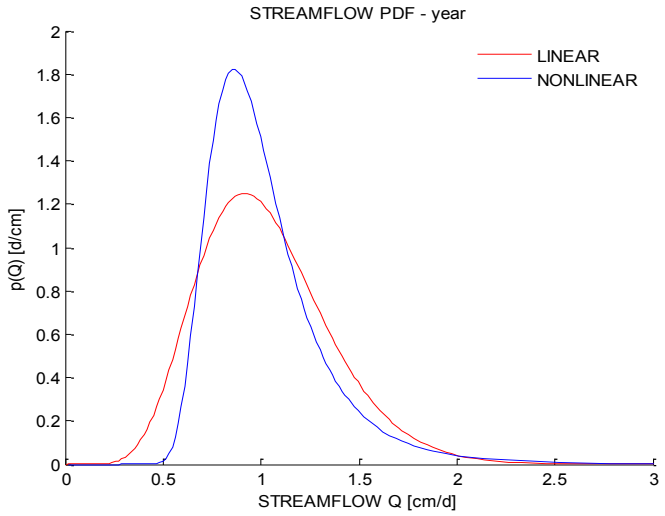


Figure 57: Streamflow pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

Figures 53 to 57 show the probability distribution of streamflow for MDD.

Streamflow is much larger in spring, the rainy season, than during the rest of the year.

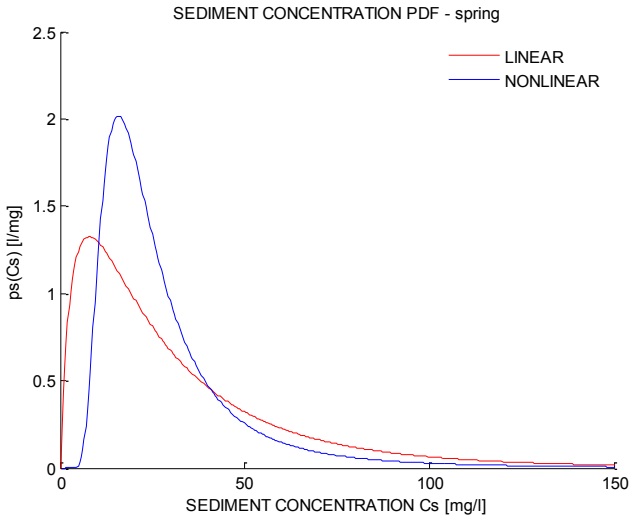


Figure 58: Sediment concentration pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

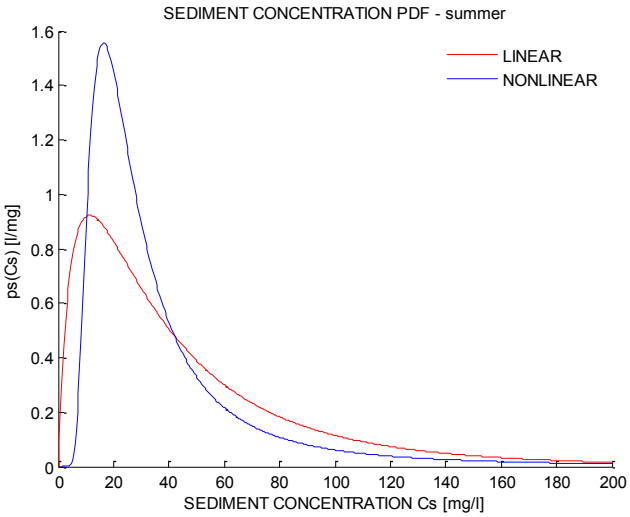


Figure 59: Sediment concentration pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

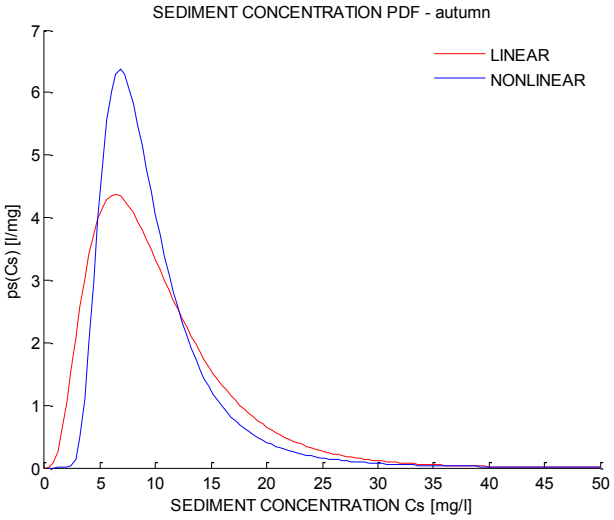


Figure 60: Sediment concentration pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

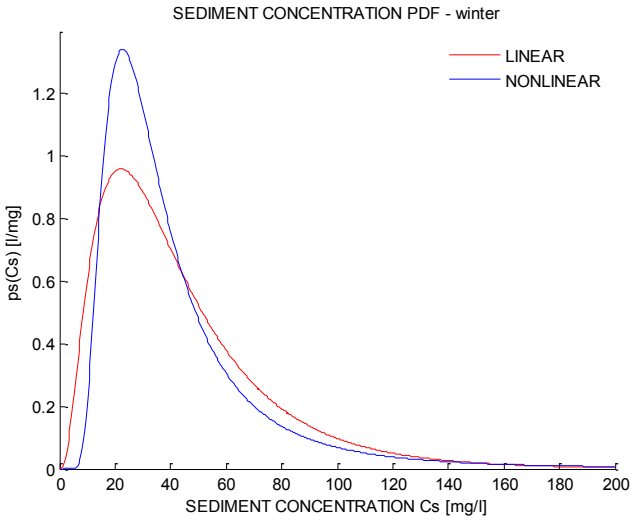


Figure 61: Sediment concentration pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

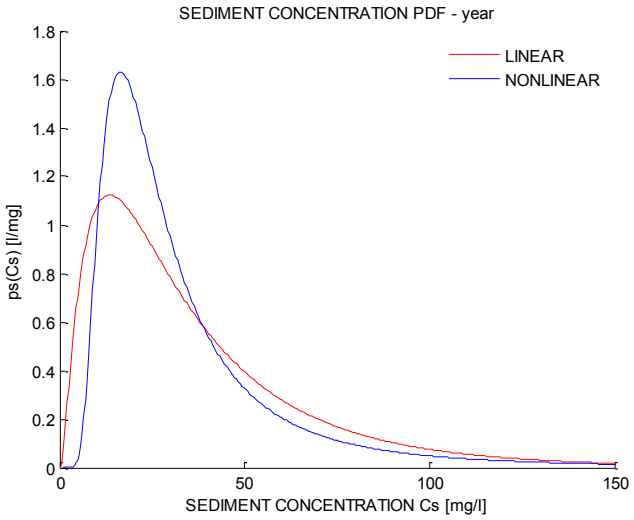
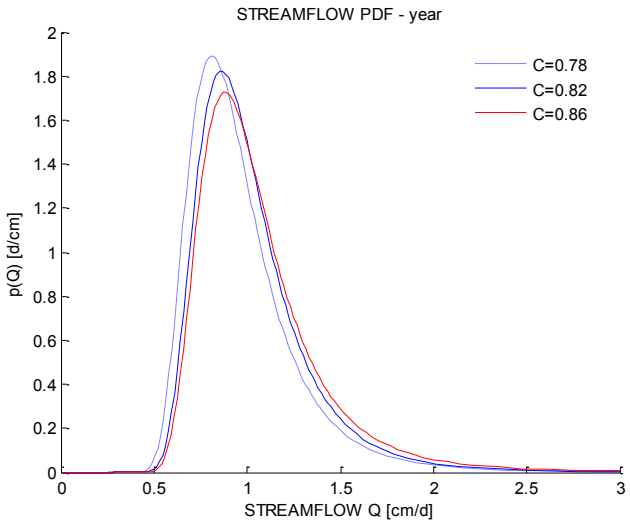


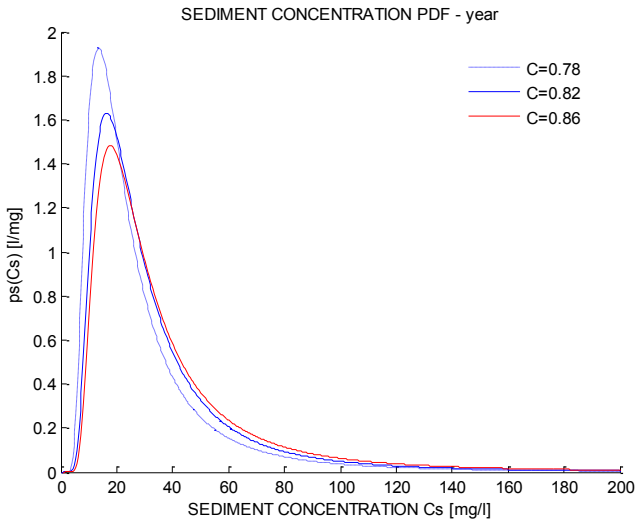
Figure 62: Sediment concentration pdf in the MDD: blue line is the nonlinear model; red line is the linear model.

Figures 58 to 62 show the probability distribution of sediment concentration for MDD.

The runoff coefficient C=0.82 is derived from the neighboring catchment Ramis (15.26°S, 69.87°W). In order to explore a range of runoff coefficient values within which the MDD runoff coefficient should realistically fall, I use here data from two more nearby catchment. CatchmentHuancane (15.22°S, 69.79°W) is characterized by a runoff coefficient C=0.78, while C= 0.86 has been estimated for the Ilave catchment (16.09°S, 69.63°W) [55]. Use of these runoff coefficient values also allows the evaluation of the uncertainty associated with inaccurate runoff coefficient assumptions.



(a)



(b)

Figure 63: Comparing results with different runoff coefficient C. a) streamflow pdf; b) sediment concentration pdf.

Figure 63 indicates that decreasing the value of the runoff coefficient C leads to decreasing mean and mode of the streamflow probability distribution, as expected.

Table 6: Range of Qs_{mean}

C	$\gamma_w(cm^{-1})$	Qs_{mean}(tons/yr)
0.78	1.1	1582
0.78	2.1	685
0.86	1.1	2332
0.86	2.1	1009

Table 6 shows the range of mean yearly sediment discharge based on maximum and minimum C and γ_w.

The results for Rio Mameyes and Rio Caonillas suggest that the nonlinear model produces more realistic results than the linear one. I thus use here the nonlinear model

to obtain an estimate of the yearly sediment discharge, the main quantity of interest in relation to the potential transport of contaminated sediment through the river. The estimate is obtained by averaging the product Qs= Q * Cs(Q) over the yearly frequency distribution, both Q and Cs(Q) are estimated from the model. The mean sediment discharge in MDD is 1500 tons/yr, which is based on yearly scale (C=0.82 γ_w=1.3 cm^{-1}). I calculate the mean seasonal sediment discharge separately. The sum of these values is 1350 tons/yr. According to the different runoff coefficient values explored, the mean sediment discharge ranges between 1224 to 1794 tons/yr. According to the sensitivity analysis, the observed seasonal range of variation of γ_w , between 1.1 and 2.1 cm^{-1}, has a smaller influence on the sediment transport estimates than the variation in runoff coefficients across neighboring catchments (between 0.78 and 0.86). Overall, the consideration of the maximum and minimum C and γ_w values in Table 6, yields a range of estimated sediment transport capacity values between 685 and 2,332 tons/yr.

Previous tests on gauged basins suggest, however, that the present method is affected by a systematic underestimation of the actual values. In fact, according to Table 5, I applied the max estimation error E=-0.205 and the minimum estimation error, E=-0.049, to obtain adjusted estimates. After correction, the mean sediment transport is from 1557 to 1887 tons/yr when C=0.82 and γ_w=1.3 cm^{-1}. The sum of seasonal transports could be from 1420 to 1698 tons/yr. The consideration of such underestimation errors thus

leads to a range of most likely values in the MDD at Puerto Maldonado between 720 to 2,933 tons/yr.

It is interesting to note that, for the same MDD study area, Berky [14] reports that "on a yearly basis a total of 2,311,451 kg of soil is expected to be eroded". Of course it should be noted that the present estimates concern the transport capacity of the river, rather than the erosion rates within the basin. The estimated transport capacity and erosion rates above are quite coherent and allow to establish the order of magnitude of sediment transport in the MDD, but do not allow to conclusively determine whether sediment supply or transport limitations occur in the MDD.

4. Conclusions

According to the above analysis some conclusions can be drawn.

Remote sensing allows to define areas which are most impacted by gold mining activities in the Madre de Dios region and to track their evolution over time. The information gathered by remote sensing can also be used to inform hydrological modeling thus supplying critical information in a data-scarce area.

The stochastic approach proposed by Botter and Basso [52] can be used to obtain first-order estimates of the frequency distributions of water and sediment fluxes using a limited number of easily determined variables. While estimation uncertainties remain high, these can be valuable exploratory tools to characterize the hydrologic cycle and sediment transport in ungauged watersheds.

The stochastic tools that have been tested on well monitored USGS study catchments have been used to provide the very first estimates of water and sediment transport in the Madre de Dios catchment. While the present application focuses on the watershed closed at Puerto Maldonado, the methods can be applied to any subcatchment in the area and thus provide useful background information to determine sources and sinks of potentially mercury-contaminated sediments.

Finally, I have applied a stochastic estimation method to determine the delivery rate of potentially contaminated sediment in the Madre de Dios region. The mean sediment discharge in the MDD is 1500 tons/yr. The overall rate of sediment yield at

Puerto Maldonado, between 720 and 2,933 tons/yr, is in good agreement with other estimates obtained using very different methodologies. Ramis catchment is closer to the MDD catchment than other two catchments Huancane and Ilave, so the runoff coefficient C should be close to 0.82. The raining season can last six month in the MDD, so the mean γ_w is around 1.3 cm^{-1}. The actual value of sediment transport is probably from 1557 to 1887 tons/yr. The model applied here thus contributes to build more accurate quantifications of the volume of potentially contaminated sediment transported into the Amazon River. The application of this method to different subcatchments, e.g. those where illegal mining is known to be most intense (e.g. through remote sensing), will allow a more spatially-detailed identification and quantification of the sediment mobilized in areas where it may be polluted by mining activities.

Appendix A

Table 7: Monthly Rainfall in Puerto Maldonado

Year	Jan.	Feb.	Mar	Apr.	May	Jun.
1998	124.6	174.5	312.3	799.6	48.7	19.3
1999	210.4	315.3	244.6	86.4	76	83
2000	290.8	336	401	75.1	66.8	115.3
2001	299.2	217.3	347.3	106.4	152.7	69
2002	117	409	178.6	215.9	166.3	163.8
2003	397.8	370.1	369.8	273.7	63.7	166.7
2004	367.5	201.2	169.5	138.8	13.6	12.7
2005	240.4	365.5	187.1	110.8	48.9	59.1
2006	603.6	164.5	218.3	405	266.7	70.1
2007	274.7	468.4	141.3	236.4	130.7	1.4
2008	468.1	296.9	192.7	44	67	0
Year	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1998	11.5	94.1	92.8	247.6	368.8	282.8
1999	18	1.7	205	74.4	197.9	194.9
2000	9.7	10.6	104.7	40.2	408.3	238.7
2001	44.5	54.5	73.6	245.4	322.5	399.7
2002	164.5	97.9	24.8	155.9	340.8	510.7
2003	2.6	119.5	151.4	347	228	373.7
2004	195.9	62.2	21.9	80	356.5	376.6
2005	25.3	14.2	25.8	156.7	231.2	339.3
2006	7.7	10.7	115.6	171	129.6	256.6
2007	102.1	8.7	49.6	152.9	254.7	284.9
2008	15.5	39.3	55.6	131.3	113.1	363.3

Table 8: Monthly Rainfall and discharge in Rio Mameyes

Time	Monthly Rainfall (mm)	Monthly Discharge (m^3/s)
Dec-92	516.5	70
Jan-93	0*	57
Feb-93	265.3	45
Mar-93	213.1	30

Apr-93	277.1	41
May-93	416.9	62
Jun-93	296.8	52
Jul-93	513.9	75
Aug-93	187.8	41
Sep-93	414.4	52
Oct-93	353.2	40
Nov-93	415.4	43
Dec-93	250.6	29
Jan-94	342.9	32
Feb-94	451.9	50
Mar-94	167	19
Apr-94	220.4	22
May-94	304.8	27
Jun-94	254.3	28
Jul-94	272.8	18
Aug-94	208.9	18
Sep-94	289.9	24
Oct-94	0*	27
Nov-94	379.5	46
Dec-94	340.3	42
Jan-95	273.9	38
Feb-95	285.6	43
Mar-95	187	32
Apr-95	0*	16
May-95	374.4	44
Jun-95	375.4	55
Jul-95	412	38
Aug-95	321	48
Sep-95	423.2	53
Oct-95	266.9	44
Nov-95	309.3	35
Dec-95	325.2	41
Jan-96	757.6	66
Feb-96	295.7	27
Mar-96	247.6	21
Apr-96	210.1	42
May-96	344.5	48
Jun-96	519.3	38

Jul-96	428.3	55
Aug-96	404.6	40
Sep-96	887.8	122
Oct-96	300.6	35
Nov-96	678.8	87
Dec-96	311.3	30
Jan-97	364.4	42
Feb-97	389.5	45
Mar-97	290.5	29
Apr-97	107.2	29
May-97	330.4	42
Jun-97	335	37
Jul-97	358	36
Aug-97	310.6	47
Sep-97	405.6	56
Oct-97	622.3	66
Nov-97	504.3	64
Dec-97	157.7	17
Jan-98	360.3	48
Feb-98	240.4	31
Mar-98	308.9	36
Apr-98	276.9	53
May-98	435.2	61
Jun-98	277.2	49
Jul-98	347.4	38
Aug-98	492.4	71
Sep-98	494.8	125
Oct-98	552.8	62
Nov-98	450.9	70
Dec-98	680.1	115
Jan-99	393.9	69
Feb-99	162.5	33
Mar-99	225.7	29
Apr-99	234.5	20
May-99	169.9	29
Jun-99	330.8	33
Jul-99	351.3	60
Aug-99	536.4	57
Sep-99	417.6	46

Oct-99	546.2	68
Nov-99	847	127
Dec-99	641.3	106
Jan-00	283.5	45
Feb-00	237.5	33
Mar-00	111.3	15
Apr-00	204.6	16
May-00	435.3	44
Jun-00	290	39
Jul-00	200.2	31
Aug-00	577.3	75
Sep-00	390.5	61
Oct-00	306	35
Nov-00	284.7	35
Dec-00	251.7	34
Jan-01	202.5	23
Feb-01	324.9	31
Mar-01	108.8	24
Apr-01	302.6	37
May-01	320.9	30
Jun-01	267.9	19
Jul-01	275.9	32
Aug-01	512.9	70
Sep-01	388.1	29
Oct-01	358.7	48
Nov-01	614.1	88
Dec-01	584.6	123
Jan-02	344.7	62
Feb-02	184.2	34
Mar-02	216.4	20
Apr-02	129.1	78
May-02	258.7	62
Jun-02	166.1	43
Jul-02	180.2	22
Aug-02	390.6	32
Sep-02	224.3	43
Oct-02	373.4	56
Nov-02	197	29
Dec-02	231.2	31

Jan-03	268.5	38
Feb-03	297.5	37
Mar-03	128.6	14
Apr-03	264	123
May-03	152.7	49
Jun-03	172.9	40
Jul-03	166.9	36
Aug-03	298.4	56

*Data is missing

Table 9: Monthly Rainfall data in Rio Caonillas

Time	Monthly Rainfall (mm)	Monthly Discharge (m^3/s)
Oct-95	280.4	142
Nov-95	91.2	78
Dec-95	52.4	41
Jan-96	106.1	44
Feb-96	114.8	73
Mar-96	42.9	45
Apr-96	202.9	50
May-96	220.8	111
Jun-96	149.7	74
Jul-96	127.1	103
Aug-96	145.4	42
Sep-96	627.6	332
Oct-96	178	64
Nov-96	279.1	100
Dec-96	51.1	69
Jan-97	149.9	79
Feb-97	21.5	39
Mar-97	101.1	23
Apr-97	17.9	15
May-97	168.6	26
Jun-97	112.7	15
Jul-97	185	14
Aug-97	204.5	55
Sep-97	301.2	29
Oct-97	262.1	106
Nov-97	96	39

Dec-97	14	18
Jan-98	68.3	17
Feb-98	159.5	41
Mar-98	98.1	37
Apr-98	166.7	98
May-98	333.5	114
Jun-98	127	67
Jul-98	168.1	55
Aug-98	351.5	137
Sep-98	288.8	0*
Oct-98	252.6	217
Nov-98	121.6	161
Dec-98	97.6	101
Jan-99	69.7	64
Feb-99	41.3	46
Mar-99	155.4	35
Apr-99	207	52
May-99	169.6	39
Jun-99	258.8	138
Jul-99	109.4	37
Aug-99	219.3	96
Sep-99	321.9	243
Oct-99	272.9	188
Nov-99	310.2	262
Dec-99	99.3	126
Jan-00	48.2	73
Feb-00	38.4	47
Mar-00	57.2	28
Apr-00	180.4	27
May-00	420.2	96
Jun-00	144.5	36
Jul-00	48.2	26
Aug-00	333.3	81
Sep-00	352.7	171
Oct-00	266.3	186
Nov-00	95.2	98
Dec-00	36.6	39
Jan-01	42.5	30
Feb-01	42.9	23

Mar-01	28.1	24
Apr-01	247.1	24
May-01	240.7	74
Jun-01	149.8	39
Jul-01	153.7	61
Aug-01	163.5	70
Sep-01	278.8	120
Oct-01	172.1	56
Nov-01	178.4	138
Dec-01	130.8	79
Jan-02	23.6	42
Feb-02	43.6	27
Mar-02	106.4	45
Apr-02	447.2	203
May-02	92.4	78
Jun-02	88.9	59
Jul-02	56.8	28
Aug-02	207.8	43
Sep-02	178.2	52
Oct-02	159.9	59
Nov-02	182.9	34
Dec-02	98.6	29
Jan-03	69.4	24
Feb-03	76.8	41
Mar-03	87	21
Apr-03	185.9	70
May-03	203.6	76
Jun-03	82.2	30
Jul-03	125.3	18
Aug-03	132.4	21
Sep-03	294.3	66
Oct-03	441.5	162
Nov-03	439.2	302
Dec-03	129.8	131
Jan-04	123.4	67
Feb-04	68.6	43
Mar-04	158.4	50
Apr-04	244.5	76
May-04	343.5	161

Jun-04	117.9	61
Jul-04	179.8	47
Aug-04	265.7	44
Sep-04	460.3	206
Oct-04	263.8	156
Nov-04	292.8	267
Dec-04	112.3	84
Jan-05	106.5	107
Feb-05	25.8	60
Mar-05	20.9	31
Apr-05	184.2	51
May-05	258.1	178
Jun-05	239.5	92
Jul-05	331.6	156
Aug-05	271	183
Sep-05	370.4	175

*Data is missing

Table 10: Column Water Vapor Amounts and Surface Temperatures for the MODTRAN Model Atmospheres (Continued).

Model Atmosphere	Water Vapor (std atm-cm)	Water Vapor (g/cm2)	Surface Air Temperature
Mid-Latitude Winter (MLW)	518	0.42	-16℃ or 3℉
Sub-Arctic Winter (SAW)	1060	0.85	-1 ℃ or 30 ℉
U.S. Standard (US)	1762	1.42	15 ℃ or 59 ℉
Sub-Arctic Summer (SAS)	2589	2.08	14 ℃ or 57 ℉
Mid-	3636	2.92	21 ℃ or 70

Latitude Summer (MLS)			℉
Tropical (T)	5119	4.11	27 ℃ or 80 ℉

Table 11: Selection of MODTRAN Model Atmospheres Based on Latitudinal/Seasonal Dependence of Surface Temperature

Latitude (°N)	Jan.	March	May	July	Sept.	Nov.
80	SAW	SAW	SAW	MLW	MLW	SAW
70	SAW	SAW	MLW	MLW	MLW	SAW
60	MLW	MLW	MLW	SAS	SAS	MLW
50	MLW	MLW	SAS	SAS	SAS	SAS
40	SAS	SAS	SAS	MLS	MLS	SAS
30	MLS	MLS	MLS	T	T	MLS
20	T	T	T	T	T	T
10	T	T	T	T	T	T
0	T	T	T	T	T	T
-10	T	T	T	T	T	T
-20	T	T	T	MLS	MLS	T
30	MLS	MLS	MLS	MLS	MLS	MLS
-40	SAS	SAS	SAS	SAS	SAS	SAS
-50	SAS	SAS	SAS	MLW	MLW	SAS
-60	MLW	MLW	MLW	MLW	MLW	MLW
-70	MLW	MLW	MLW	MLW	MLW	MLW
-80	MLW	MLW	MLW	SAW	MLW	MLW

Table 12: Initial visibility value.

Weather Condition	Scene Visibility
Clear	40 to 100 km
Moderate Haze	20 to 30 km
Thick Haze	15 km or less

References

[1] Larmer B (2009) The real price of Gold. Nat Geo. Jan 2009.

[2] Keane L (2009) Rising prices spark a new gold rush in Peruvian Amazon. 19 December. Washington DC, U S A: Washington Post, Available: http://www.washingtonpost.com/wpdyn/content/article/2009/12/18/AR2009121804139.html. Accessed 2010 Jan 10.

[3] Food and Agriculture Organization (FAO). (2005; 2010). Global Forest Resources Assessments (FRA).http://www.fao.org/forestry/fra/en/

[4] Brooks WE, Sandoval E, Yepez MA, Howell H (2007) Peru Mercury Inventory 2006: U.S. Geological Survey Open-File Report 2007-1252. Available http://pubs.usgs.gov/of/2007/1252/. Accessed 2009 Dec 15. 55 p

[5] Veiga MM, Maxson PA, Hylander LD (2006) Origin and Consumption of Mercury in Small-Scale Gold Mining. J Cleaner Prod 14: 436–447

[6] Elie Gardner, Peru battles the golden curse of Madre de Dios Attempts to reduce the environmental and health impacts of mining cause unrest. Nature 486, 306–307 (21 June 2012

[7] Velga et al, 1994 , in Kricher, 1997, Extracting valuable minerals and a Pandora's Box of problems

[8] Botter, G., A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2007a), Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: low, leaching-prone components of the hydrologic response, Water Resour. Res., 43, W02417, doi:10.1029/2006WR005043.

[9] Botter, G., A. Porporato, I. Rodriguez - Iturbe, and A. Rinaldo (2009), Nonlinear storage - discharge relations and catchment streamflow regimes, Water Resour. Res., 45, W10427, doi:10.1029/2008WR007658.

[10] Evaluacion Ambiental Territorial en la Zona Aurifera de Madre de Dios, 4. Climatología E Hidrologia Ministry of Energy and Mines, Peruvian Government, December 1997.

[11] Harmonized World Soil Database, http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/HWSD_Data.html?sb=4

[12] www.regionmadrededios.gob.pe

[13] Autoridad Nacional del Agua – Dirección de Conservación y Planeamiento de Recursos Hídricos, Diagnostico y Plan de Gestión de los Recursos Hídricos en la Cuenca de Madre de Dios – Fase I

[14] Axel Berky, Mercury Contamination of Sediment from Artisanal Gold Mining in Madre de Dios, Peru May 2014, unpublished.

[15] The World Factbook – Puerto Rico Geography. Cia.gov. Retrieved October 30, 2011.

[16] Welcome to Puerto Rico! topuertorico.org. Retrieved December 30, 2007

[17] Los Lagos de Puerto Rico at the Wayback Machine (archived June 29, 2007) (archived from the original on June 29, 2007).

[18] Andrzej Pisera, Michael Martínez, Hernan Santos (May 2006). "Late Cretaceous Siliceous Sponges From El Rayo Formation, Puerto Rico". Journal of Paleontology. Retrieved May 6, 2008.

[19] NOAA Online Weather Data – Puerto Rico. National Weather Service. Retrieved May 6, 2008.

[20] Edward B. Rodgers, Robert F. Adler, Harold F. Pierce. Contribution of Tropical Cyclones to the North Atlantic Climatological Rainfall as Observed from Satellites.

[21] USGS Caribbean Water Science Center http://maps.waterdata.usgs.gov/mapper/index.html

[22] Landsat Project Description, http://landsat.usgs.gov/about_project_descriptions.php

[23] Landsat Thematic Mapper (TM), https://lta.cr.usgs.gov/TM

[24] Landsat Processing Details, http://landsat.usgs.gov/Landsat_Processing_Details.php

[25] Ceola, S., G. Botter, E. Bertuzzo, A. Porporato, I. Rodriguez - Iturbe, and A. Rinaldo (2010), Comparative study of ecohydrological streamflow probability distributions,Water Resour. Res., 46, W09502, doi:10.1029/2010WR009102.

[26] Radiometric Correction of Remotely Sensed Data, Douglas A. Stow

[27] Atmospheric Correction. University of Maryland Institute for Advanced Computer Studies. Archived from the original on 7 September 2008. Retrieved 2008-08-18.

[28] Atmospheric Correction: Empirical Line Fits, http://www.oceanopticsbook.info/view/remote_sensing/level_2/atmospheric_correction_empirical_line_fits

[29] Anderson et al., MODTRAN4: radiative transfer modeling for remote sensing, Proc. SPIE 3866, Optics in Atmospheric Propagation and Adaptive Systems III, 2 (December 10, 1999); doi:10.1117/12.371318

[30] P.Gessler, http://www.cnr.uidaho.edu

[31] Atmospheric Correction Module: QUAC and FLAASH User's Guide, Atmospheric Correction Module Version 4.7 August, 2009 Edition

[32] Kaufman, Y. J., A. E. Wald, L. A. Remer, B.-C. Gao, R.-R. Li, and L. Flynn, 1997. The MODIS 2.1-m Channel-Correlation withVisible Reflectance for Use in Remote Sensing of Aerosol. IEEE Transactions on Geoscience and Remote Sensing. Vol. 35, pp. 1286-1298.

[33] Using ENVI and Geographic Information Systems (GIS) Thursday, January 31, 2013

[34] Tou, J. T. and R. C. Gonzalez, 1974. Pattern Recognition Principles, Addison-Wesley Publishing Company, Reading, Massachusetts.

[35] Exelis VIS product documentation center, http://www.exelisvis.com/docs/KMeansClassification.html

[36] Steinhaus, H. (1957). "Sur la division des corps matériels en parties". Bull. Acad. Polon. Sci. (in French) 4 (12): 801–804. MR 0090073. Zbl 0079.16403.

[37] MacKay, David (2003). "Chapter 20. An Example Inference Task: Clustering". Information Theory, Inference and Learning Algorithms. Cambridge University Press. pp. 284–292. ISBN 0-521-64298-1. MR 2012999.

[38] Geoffrey H. Ball and David J. Hall, ISODATA, a novel method of data analysis and pattern classification, Stanford Research Institute, 1965.

[39] ISODATA, a novel method of data analysis and pattern classification, Stanford Research Institute, 1965.

[40] Unsupervised Classification – Clustering, Ruye Wang 2013-12-09 http://fourier.eng.hmc.edu/e161/lectures/classification/node12.html

[41] IMAGE PROCESSING AND INTERPRETATION: MORRO BAY, CALIFORNIA, Nicholas M. Short, Sr. Jon Robinson, 1999

[42] Center for the Study of Earth from Space (CSES), 1992, SIPS User's Guide, The Spectral Image Processing System, v. 1.1, University of Colorado, Boulder,74 p.

[43] Kruse, F. A., A. B. Lefkoff, J. B. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, 1993, "The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging spectrometer Data." Remote Sensing of the Environment, v. 44, p. 145 - 163.

[44] Swenson JJ, Carter CE, Domec J-C, Delgado CI (2011) Gold Mining in the Peruvian Amazon: Global Prices, Deforestation, and Mercury Imports. PloS ONE 6(4): e18875. doi:10.1371/journal.pone.0018875

[45] Gregory P. Asner, William Llactayo, Raul Tupayachi, and Ernesto Ráez Luna (2013), Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring, PNAS Early Edition, www.pnas.org/cgi/doi/10.1073/pnas.1318271110

[46] Porporato, A., E. Daly, and I. Rodriguez-Iturbe (2004), Soil water balance and ecosystem response to climate change,Am. Nat., 164(5), 625 – 632.

[47] Rodriguez-Iturbe, I., and A. Porporato (2004), Ecohydrology of WaterControlled Ecosystems: Soil Moisture and Plant Dynamics, Cambridge Univ. Press, New York.

[48] Milly, P. C. D. (1993), An analytic solution of the stochastic storage problem applicable to soil water, Water Resour. Res., 29(11), 3755 – 3758.

[49] Botter, G., F. Peratoner, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2007c), Signatures of large-scale soil moisture dynamics on streamflow statistics across U.S. climate regimes, Water Resour. Res., 43, W11413, doi:10.1029/2007WR006162.

[50] James P. Syvitski, Mark D. Morehead, David B. Bahr, and Thierry Mulder, Estimating fluvial sediment transport: The rating parameters, WATER RESOURCES RESEARCH, VOL. 36, NO. 9, PAGES 2747-2760, SEPTEMBER 2000.

[51] Brutsaert, W., and J. L. Nieber (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, Water Resour. Res., 13(3), 637–648.

[52] Basso and Botter, in preparation, 2014.

[53] Botter G, Basso S, Rodriguez-Iturbe I, Rinaldo A (2013), Reslilience of river flow regimes.PNAS August 6, 2013 vol. 110 no. 32 12925–12930.

[54] Mikdat Kadioglu, Zekai Sen, Monthlyprecipitation-runoff polygons and mean

Runoffcoefficients Hydrological Sciences-Jourtial-des Sciences Hydrologiques, 46(1)Februar 2001

[55] Waldo Sven Lavado Casimiro, Josyane Ronchail, David Labat, Jhan Carlo Espinoza & Jean Loup Guyot (2012): Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages, Hydrological Sciences Journal, DOI:10.1080/02626667.2012.672985

[56] Syvitski, J.P. M., D.C. Burrell, and J. M. Skei, Fjords: Processes and Products, 379 pp., Springer-Verlag, New York, 1987.